Early language variation and working memory: A longitudinal study of late talkers and typically developing children

A thesis submitted for the degree of Doctor of Philosophy at the University of Canterbury

2014

Jayne Margaret Newbury (née Moyle)

Department of Communication Disorders

University of Canterbury

Abstract

This research explored whether variation in working memory ability helps account for the wide variation in toddlers' language skills and improves predictive models of language outcomes over time. A cohort of typically developing (TD) (*n* = 55) and late talking children (*n* = 24) were assessed at two time points. The initial assessment took place at ages 24-30 months and the outcome assessment occurred 18 months later, when the children were aged 41-49 months. The assessment battery included standardised tests of language and visual cognition; assessments representing aspects of Baddeley's model of working memory: phonological short term memory (PSTM), a measure of processing speed, verbal working memory (VWM), visual spatial working memory (VSWM), and a parent report questionnaire of executive functioning (EF). Study 1 explored the associations between these aspects of working memory and concurrent expressive vocabulary at ages 24-30 months and examined group differences in the measures between TD and late talking children. Study 2 explored associations between aspects of working memory and concurrent expressive language in the same cohort at 41-49 months of age. Group differences in the measures between resolved late talkers (RLTs) and TD children were explored. Finally Study 3 explored the ability of the measures used at 24-30 months to predict language outcomes at 41-49 months. These results were considered in relation to the prediction of language outcomes on group and individual levels. Overall the results indicated a strong relationship between early PSTM and early language measures. A novel finding was that PSTM was significantly lower in the late talking and RLT groups compared with the TD groups, even after controlling for group differences in language and phonology at both time points. This confirms previous research that PSTM plays a role in early expressive vocabulary acquisition, and suggests that early PSTM deficits may be a causal factor for some cases of late talking. For the whole group, three working memory variables (VWM, Emotional Control and Shift) measured at 24-30

months added unique variance to predictive models in total language scores at 41-49 months after previously established early predictors (receptive language and parent education) had been entered into the hierarchical regression model (receptive language $R^2\Delta$ = 59%; parent education $R^2\Delta$ = 2%; VWM $R^2\Delta$ = 8%; Emotional Control $R^2\Delta$ = 1% and Shift $R^2\Delta$ = 2%). This is another novel finding which supports the concept of working memory playing a unique role in language acquisition between the ages two and four years. Processing speed did not contribute unique variance to regression models predicting language when other working memory measures were included. The A not B task (measuring VSWM) did not correlate with language. There were concerns with construct validity with the EF parent report measure (Behaviour Rating Inventory of Executive Function – Preschool Version), which meant that the results from this assessment were interpreted with caution. In terms of clinical outcomes, 83% of the late talkers resolved their language delays over the 18 month period, but as a group showed a seven-fold increase in being identified for clinical concerns at the outcome assessment than children who were not late talkers. The majority of these concerns were for poor phonology. While early VWM, Shift and Emotional Control added unique variance to outcome total language scores on a group level, they did not improve prediction of individual outcomes in language impairment status at 41-49 months. Early receptive language delay was a more powerful predictor of later language impairment than late talking in this cohort, as these children (n = 9) showed only a 44% rate of resolution.

Acknowledgements

Special thanks go to my supervisors, Drs Tom Klee, Stephanie Stokes and Catherine Moran. In particular, I have appreciated the hours of listening and questioning provided by Tom. His patient reframing of my thoughts by challenging my many assumptions has trained me out of thinking in a purely clinically framework and made me into an academic.

My study was run as a part of a wider research project ("Learning to Talk") which was funded by the Marsden Fund of the Royal Society of New Zealand. My PhD studies were funded by a doctoral scholarship from the University of Canterbury and also through the Marsden grant. Without these funding sources, this research would not have been possible.

Thanks to my friends and family for encouraging me that I could do this right at the beginning, and for encouraging me to keep going right to the end!

Thanks to everyone who helped to recruit participants, in particular to the early childhood centres in Christchurch who spread the word (especially Kidicorp), Pegasus Health, the Champion Centre, the Ministry of Education (Special Education), and Plunket.

Thank you to all the children who were patient with the many tasks they were asked to do (even when they would rather be outside playing), and thanks also to their parents for their commitment to this project!

Thanks to Drs William Gavin and Margaret McClaghan and also to Sarah Kerr who provided invaluable technical support for the Looking While Listening I task. Thanks also to Dr Elena Molchanova, my statistical advisor, for her patient explanations.

Thank you to the research assistants working on the Learning to Talk project who helped with the many hours of coding and scoring in Study 1: Jenna McAffee, Jane McKinnon and Asifa Sultana, and also to our honours student Emma Irvine.

Finally I acknowledge my loving God as creator and guide throughout this journey. Thanks for making the way clear for me.

For my husband Garrett, whose generosity of heart, love and support inspires me daily to work harder to make a difference for children with learning difficulties.

Table of Contents

Abstract *ii*

Acknowledgements *iv*

Dedication *v*

Contents *vi*

List of Figures *xii*

List of Tables *xiii*

List of Abbreviations *xv*

Chapter One: Predicting outcomes for late talkers

1.0 Aims of this research 2

1.1 Late talkers 3

1.1.1 Late talker cohorts 4

1.1.2 Developmental factors as predictors of language outcomes 7

1.1.3 Environmental factors as predictors of language outcomes 10

1.1.4 Genetic factors as predictors of language outcomes 12

1.1.5 Multiple risk factor models 14

1.2 Theories of specific language impairment (SLI) 17

Chapter Two: Working memory and language acquisition

2.1 Introduction to working memory 20

2.2 Baddeley and Hitch's model of working memory 21

2.3 Phonological short term memory 23

2.3.1 The role of phonological short term memory (PSTM) in language acquisition 26
2.4 Declarative-procedural memory systems 28
2.5 Processing speed 31
2.6 Central executive (CE) 32
2.6.1 Verbal working memory (VWM) 33
2.6.1.1 The role of verbal working memory (VWM) in language processing 35
2.6.1.2 Evidence linking verbal working memory with language acquisition 37
2.6.2 Visual-spatial working memory (VSWM) 40
2.6.3 Executive functions (EF) 41
2.6.3.1 The relationship between executive functions and language acquisition 43
2.6.3.2 Inhibition 43
2.6.3.3 Shifting 46
2.6.3.4 Metacognitive executive functions 47
2.6.3.5 Plan / Organise 48
2.6.3.6 Emotional Control 49
2.7 Summary of literature review 50
2.8 Overview of the current research 51

Chapter Three: Study 1 Methodology

3.1 Recruitment 56
3.2 Participants 58

3.3 Measures 62

3.3.1 Questionnaires 62

3.3.2 In-clinic behavioural measures 63

3.4 Procedures 77

Chapter Four: Study 1 Results

4.1 Missing data analysis 80

4.2 Descriptive statistics for the assessment data 81

4.3 Research question 1 83

4.3.1 Bivariate associations between the variables 83

4.3.2 Effect of demographic variables on expressive vocabulary 85

4.3.3 Initial multivariate regression model for expressive vocabulary at 24-30 months 88

4.4 Research question 2 90

4.4.1 Comparison multivariate regression model for expressive vocabulary 90

4.5 Research question 3 93

4.5.1 Criteria for late talking 93

4.5.2 Comparison of group means 94

4.6 Research question 4 98

4.6.1 Comparison of group means in working memory while controlling for language 98

4.7 Chapter summary 100

Chapter Five: Study 2 Methodology

5.1 Inviting participants to return 104

5.2 Descriptive statistics for demographic variables 105

5.3 Changes to the measures 106

5.4 Procedures 112

Chapter Six: Study 2 Results

6.1 Missing data analysis 115

6.2 Descriptive statistics for the assessment data 116

6.3 Research question 5 118

6.3.1 Bivariate associations between the concurrent variables 118

6.3.2. Effect of demographic variables on expressive language at 41-49 months 120

6.3.3 Initial multivariate regression model for concurrent expressive language at 41-49 months 123

6.4 Research question 6 125

6.4.1 Comparison multivariate regression model for concurrent expressive language at 41-49 months 126

6.5 Research question 7 128

6.5.1 Comparison of group means at 41-49 months between RLT and TD groups 128

6.6 Research question 8 133

6.6.1 Comparison of group means in working memory while controlling for language at 41-49 months 134

6.7 Chapter summary 136

Chapter Seven: Study 3 Results

7.1 Research question 9 139

7.1.1 Clinical outcomes at 41-49 months of age 139

7.2 Research question 10 143

7.2.1 Bivariate associations between the measures at 24-30 months and PLS-4 total language scores at 41-49 months 144

7.2.2 Initial multivariate regression model for total language scores at 41-49 months using variable measured at 24-30 months 145

7.3 Research question 11 149

7.3.1 Comparison multivariate regression model for total language outcomes at 41-49 months using variables measured at 24-30 months 149

7.4 Research question 12 152

7.4.1 Individual prediction 152

7.5 Chapter summary 154

Chapter Eight: General discussion

8.1 Introduction 157

8.2 Phonological Short Term Memory (PSTM) 158

8.3 Processing speed 164

8.4 Summary of findings for the Central Executive (CE) 166

8.4.1 Verbal working memory (VWM) 166

8.4.2 Visual spatial working memory (VSWM) 169

8.4.3 Executive functions (EF) 171

8.4.3.1 Behaviour Rating Inventory of Executive Functions – Preschool Version (BRIEF-P) 171

8.4.3.2 Shift 172

8.4.3.3 Emotional Control 173

8.4.3.4 Inhibition 174

8.4.3.5 Working Memory 175

8.4.3.6 Plan / Organise 176

8.4.3.7 Summary of Executive Functions (EF) and language 177

8.5 Models of individual prediction 178

8.6 Limitations of the current research 181

8.7 Strengths of the current research 181

8.8 Implications for clinical practise 182

8.9 Future research 183

8.10 Conclusion 183

Appendix A: Ethics approval 185

Appendix B: Parent pack for Part 1 186

Appendix C: Parent pack for Part 2 201

Appendix D: Score forms 204

Appendix E: Looking While Listening Task I coding procedures 214

Appendix F: Diagnostic procedures for the three multiple linear regression Models 220

Appendix G: Parent Questionnaire II 225

References 226

List of Figures

Figure 2.1	Baddeley's model of working memory	22
Figure 2.2	Cowan's model of working memory	24

List of Tables

Table 1.1	Comparison of the latest four multiple risk factor models for ongoing delays in late talkers	14
Table 3.1	Summary of parent and child demographic characteristics when the children were aged 24-30 months	59
Table 3.2	Comparison of the sample and population ethnic mix	60
Table 3.3	Highest parent qualifications in the sample compared with population frequencies	61
Table 4.1	Percentage of missing data by assessment at 24-30 months	80
Table 4.2	Descriptive statistics for the assessment data at 24-30 months	82
Table 4.3	Bivariate correlations between the variables at 24-30 months (N = 79)	84
Table 4.4	Multiple regression analysis summary for concurrent variables predicting expressive vocabulary (CDI scores) at 24-30 months (N = 77)	89
Table 4.5	Hierarchical multiple regression analysis summary for concurrent variables predicting expressive vocabulary (CDI scores) at 24-30 months (N = 77)	91
Table 4.6	Comparison of late talker and TD group results for the behavioural assessments	96
Table 5.1	Summary of child characteristics at ages 41-49 months	105
Table 5.2	Pilot study results for the comparison of two versions of the A not B task	109
Table 6.1	Number and percentage of missing data for each measure (N = 78)	115
Table 6.2	Descriptive statistics for the behavioural assessments at outcome (aged 41-49 months)	116
Table 6.3	Bivariate correlations between the variables at 41-49 months (N = 78)	119

Table 6.4	Multiple regression analysis summary for concurrent variables predicting expressive language (PLS-4 EC raw scores) at 41-49 months (N = 74)	124
Table 6.5	Hierarchical multiple regression analysis summary for concurrent variables predicting expressive language (PLS-4 EC raw scores) at 41-49 months (N = 74)	127
Table 6.6	Comparison of RLT, TD and LI groups results for the behavioural assessments	131
Table 7.1	Summary of clinical outcomes at 41-49 months for the whole group, LT and TD groups (groups defined at ages 24-30 months)	140
Table 7.2	Bivariate partial correlations between PLS-4 total standard scores at 41-49 months and the behavioural assessments at 24-30 months	145
Table 7.3	Multiple regression analysis summary for variables measured at 24-30 months predicting PLS-4 total language standard scores at 41-49 months (N = 76)	146
Table 7.4	Hierarchical multiple regression analysis summary for variables measured at 24-30 months predicting PLS-4 total language standard scores at 41-49 months (N = 76)	150

List of abbreviations

ADHD……………………………………..……..Attention Deficit Hyperactivity Disorder

BRIEF-P…………Behavioural Rating Inventory of Executive Functions – Preschool version

CAS…………………………………….……………………Childhood Apraxia of Speech

CDI…………………..……..….MacArthur Bates Communicative Development Inventory

CE ……………………………………………………….………………… Central Executive

DEAP……………………..…………Diagnostic Evaluation of Articulation and Phonology

EF ……………………………………….……………….……………… Executive Function

fMRI ………………………..…….……….…. Functional Magnetic Resonance Imaging

GEC …………………………….………………………… Global Executive Composite

KWL…………………………….……………………………………Key Word Level

KWM………………………………..………………..Key Word Working Memory Task

LI..…………………………………….………….………………. Language Impairment

LTM ……………………………….……………………………. Long Term Memory

LWL task……………………………………….……………Looking While Listening task

NCEA …………………….….………. National Certification of Educational Achievement

NWR ………………………..…..…………………………………….. Non-word repetition

NZ ……………………………………..……………….…………………….. New Zealand

NZQA ……….…………………………….………………… New Zealand Qualifications Authority

OAE……………………………………………………………………Otoacoustic Emissions

PCC………………...…………………………………….Percentage Consonants Correct

PEDS ……………………..…….……………… Parent Evaluation of Developmental Status

PLS-4 AC…………… Preschool Language Scales (Fourth Edition) Auditory Comprehension

PLS-4 EC………….Preschool Language Scales (Fourth Edition) Expressive Communication

PSTM …………………………..……………………… Phonological Short Term Memory

RLT……………………………………………..…………..……Resolved Late Talker

ROWPVT-4....……………….Receptive One Word Picture Vocabulary Test (Fourth Edition)

SES………………………………………..……………………Socioeconomic Status

SLI ……………………….…………………….……………….. Specific Language Impairment

STM ……………….……………………….……………….………………… Short Term Memory

TD ………………………………………………….……………………….. Typically Developing

TENR………………………………...………………Test of Early Non-word Repetition

TPT…………………………...…………………………………Toddler Phonology Test

VRO ……………………………….…...…………….. Visual Reception Organisation subtest

VSWM ……………………………………..……………. Visual Spatial Working Memory

VWM ……………………………………….…………………... Verbal Working Memory

Chapter One: Predicting outcomes for late talkers

1.0 Aims of this research

1.1 Late talkers

1.1.1 Late talker cohorts

1.1.2 Developmental factors as predictors of language outcomes

1.1.3 Environmental factors as predictors of language outcomes

1.1.4 Genetic factors as predictors of language outcomes

1.1.5 Multiple risk factor models

1.2 Theories of specific language impairment (SLI)

1.0 Aims of this research

The practical question which drives this research is: “Is there anything wrong with my late talking toddler?” Well-meaning friends and relatives may say that late talkers are just “late bloomers”, and many parents take a “wait and see” approach. While this works well for children who catch up quickly, late talking can be the first observable indicator that a child has a language impairment (LI) (Buschmann et al., 2008). Children with LI benefit from early identification and intervention. Distinguishing between these two groups of children is problematic. Many studies which aimed to improve predictive models of language outcomes for late talkers have been published, focusing on a range of developmental, environmental and genetic predictors. To date, the best predictive models have been moderately successful on a group level, but largely unsuccessful at predicting individual outcomes. We have suggested that improving predictive models requires investigating a relatively unexplored area in late talker research, that is, psycholinguistic processing deficits (Moyle, Stokes, & Klee, 2011). This claim is based on a review of the evidence supporting the main causal theories of language impairments and the lack of definitive success with group and individual prediction to date. In particular, the role of working memory in early language acquisition is not well understood. To date the aspects of the working memory system which are relevant for language (the central executive (CE), phonological loop and processing speed) have not been researched together in a longitudinal cohort of late talking and typically developing (TD) two year old children. The overall aim of this research is to explore the general hypothesis that working memory measures will improve the accuracy of predictive models of language outcomes in TD and late talking toddlers in the preschool years. This research will contribute to the knowledge base supporting clinical decisions regarding the management of late talking toddlers.

1.1 Late talkers

"Late talkers" are typically identified in the age range 18-30 months. There is a wide variation in language skills in two year olds. To illustrate this, the MacArthur-Bates Communication Development Inventory (CDI) normative data indicates that the 90th percentile for expressive vocabulary for 24 month olds is 542 words, whereas the 10th percentile is only 77 words (Fenson et al., 2007). There are a range of cut-offs used to identify late talkers, from fewer than 50 expressive words on the Language Development Survey (Paul, 1996; Rescorla, 1989); and / or no two word combinations at two years of age (Klee et al., 1998); to those falling more than one standard deviation below the mean on the Communication section of the Ages and Stages Questionnaire (Bricker & Squires, 1999; Zubrick, Taylor, Rice, & Slegers, 2007); a six month delay in expressive language on the Reynell Developmental Language Scales (Rescorla, 2000; Reynell, 1977); or cut-offs at the 10th, 15th or 20th percentiles for expressive vocabulary on the MacArthur Bates Communicative Development Inventory (Beckage, Smith, & Hills, 2011; Reilly et al., 2007; Thal, Reilly, Seibert, Jeffries, & Fenson, 2004). Late talker studies may or may not include children with concomitant delays in receptive language or visual cognition. Therefore the common denominator in this group is the delay in expressive language and heterogeneity in other aspects of development is expected.

Regardless of which cut-off for expressive language is used, the majority of late talkers will resolve their initial delay and fall within normal range by ages three to four years (Dale, Price, Bishop, & Plomin, 2003; Paul, 1996; Rescorla & Roberts, 1997; Whitehurst & Fischel, 1994). The children who resolve are typically referred to as "late bloomers". This group tends to achieve in the low average range in literacy and language throughout their schooling, and may show some subclinical weaknesses in verbal working memory skills (Paul, Hernandez, Taylor, & Johnson, 1996; Preston et al., 2010; Rescorla, 2009). The variability of early language development combined with the pattern of resolution which many children display at age five years, means that the distinction between an early language *delay* and a language *impairment* is a difficult one to make in the early preschool years.

LIs are relatively common childhood disorders which have the potential to have adverse effects on children's outcomes. There are several ways LIs can be defined. Specific Language Impairment (SLI) is defined as a score of less than 1-1.5 standard deviations below the mean on one or more subtests of a standardised language measure, in the absence of sensory, environmental, cognitive or social emotional difficulties (Leonard, 1998). When other developmental difficulties are present, the LI is referred to as a non-specific LI. Prevalence of SLI was estimated at 7.4% of five year olds in the USA (Tomblin et al., 1997). Non-specific LI has been estimated at a prevalence of 12.6% of five year olds in Canada (Beitchman, Nair, Clegg, & Patel, 1986). Children whose language is in the impaired range at ages five to six years may experience poor outcomes in terms of academic achievement throughout the school years, (Stothard, Snowling, Bishop, Chipchase, & Kaplan, 1998) and are likely to continue to have a range of adverse effects such as limited vocational options, lower socio-economic status (SES) and continued depressed scores in language, cognitive and literacy measures in adulthood (Clegg, Hollis, Mawhood, & Rutter, 2005). Early identification and intervention for such children is ideal, as it maximises the amount of time the child is in an optimal language learning environment, which should improve outcomes by giving the child more opportunities to learn.

1.1.1 Late talker cohorts

While there has been a large body of research on late talkers over the last 20 years, a limitation on drawing conclusions has been the differing exclusion criteria used across studies. This section outlines evidence to support the view that a broad inclusion of participants in longitudinal studies will be maximally informative for early identification of

LI. Much of the early published information on the communicative characteristics of late talkers and their outcomes is based on four small scale middle-class American longitudinal studies which began in the 1990s: the Portland sample (Paul, 1996); Pennsylvania cohort (Rescorla & Roberts, 1997); New York cohort (Whitehurst & Fischel, 1994) and the San Diego sample (Thal, Miller, Carlson, & Vega, 2005). The resolution rates from these early late talker studies were high (for example 74% -100% by five years of age (depending on the outcome measure) in the Portland sample (Paul, 1996)), and thus fail to account for the prevalence rates of LI in the five-year-old population. These initial late talker studies excluded children with low non-verbal IQ (score below 85 on standardised testing). This could partially account for the high rates of resolution. A clear dichotomy is made in terminology between those with SLI and those with a general developmental delay; yet in reality, some children sit close to the border between these two groups and many change group membership over time (Vig, Kaminer, & Jedrysek, 1987; Webster, Majnemer, Platt, & Shevell, 2004). In addition, the exclusion of children later diagnosed with other neurodevelopmental disorders can be questioned. Boundaries between LI and other neurodevelopmental disorders are also not as distinct as they may appear. For example an overlap between dyslexia and SLI has been postulated with the shared variance thought to be due to an underlying phonological processing disorder (Catts, Adolf, Hogan, & Ellis Weismer, 2005; Schuchardt, Bockmann, Bornemann, & Maehler, 2013). The categories of autism and SLI may overlap to some extent because of a shared underlying auditory perceptual processing difficulty (Oram Cardy, Flagg, Roberts, & Roberts, 2008). There is also a moderate rate of comorbidity with SLI and Attention Deficit Hyperactivity Disorder (ADHD); both groups have difficulties with attention and working memory (Beitchman, Nair, Clegg, Ferguson, & Patel, 1986). Diagnosticians attempt to identify clusters of children and create a diagnostic category to describe them; however, there are always those who sit at the boundaries between types. Late talkers have been later diagnosed with SLI, dyslexia, ADHD, autism or general developmental delay, moving between diagnostic categories as they mature (Buschmann et al., 2008). Studies have also identified children who were typically developing at two years of age, yet fell into the impaired range over time (Dale et al., 1998; Poll & Miller, 2013; Rice, Taylor, & Zubrick, 2008). In light of the above, including late talkers who have delays in areas other than expressive vocabulary will cover a more representative sample of those who are later diagnosed with LIs and therefore be more informative. Finally, as children with SLI, resolved late talkers (RLTs) and TD children appear to represent a spectrum of language ability (Rescorla, 2009; Tomblin & Zhang, 1999), studying factors affecting typical language development would also be informative, and will be considered in this review. In contrast, children whose LI has arisen from a single mutation of a gene (e.g. Fragile X) or a physical abnormality impacting on communication skills (e.g. cerebral palsy, cleft palate or permanent hearing loss) are considered to be different populations, and studies focusing on these children are excluded from this review.

This review extends previous reviews, in that literature covering a large number of children in a range of different communities and populations, has been consulted. Other than the early American late talker studies, large scale cohorts representing a wide cross-section of communities have been published in Australia (Reilly et al., 2007; Zubrick et al., 2007); New Zealand (Silva, Williams, & McGee, 1987); the United Kingdom (Tomblin, Hardy, & Hein, 1991) and Europe (de Koning, de Ridder-Sluiter, van Agt, & Reep-van den Bergh, 2004; Henrichs et al., 2011; Maatta, Laakso, Tolvanen, Ahonim, & Ara, 2012; Westerlund, Berglund, & Eriksson, 2006; Zambrana, Ystrom, Schjølberg, & Pons, 2013). In addition, some studies focussing on specific populations have been published, such as clinical samples (Chiat & Roy, 2008); lower SES families (La Paro, Justice, Skibbe, & Pianta, 2004); children with a family history of dyslexia (Lyytinen, Poikkeus, Laakso, Eklund, & Lyytinen, 2001);

twins (Dale et al., 2003) and low birth-weight and premature children (Lars Smith & Ulvund, 1998). These studies and others have been examined for patterns of success and failure in building predictive models for language outcomes in late talking and TD toddlers. Across these studies, researchers have focused on predictive factors in three main categories: developmental, environmental and genetic factors. The main points from this body of research in these three areas are summarised below.

1.1.2 Developmental factors as predictors of language outcomes

A range of developmental variables have been studied as predictors of later language outcomes, including expressive language, speech production, receptive language, non-verbal IQ, non-verbal communication and social skills. A brief summary is presented here to capture the general pattern of findings.

There has been low to moderate success in using toddlers' early expressive language skills as a predictor of later language outcomes. Delayed expressive vocabulary measured at 18-24 months alone is not a strong predictor of language abilities two to three years later (Dale et al., 2003; Paul et al., 1996; Rice et al., 2008), although early growth patterns in expressive vocabulary may prove to be more useful in predicting later outcomes, compared with static measurement (Rowe, Raudenbush, & Goldin-Meadow, 2012).

Early speech production and oro-motor skills have had some success as predictors of later language. Early phonological development has been found to predict short term language outcomes in some studies (Thal, Oroz, & McCaw, 1995; Whitehurst, Smith, Fischel, Arnold, & Lonigan, 1991; Williams & Elbert, 2003) however, other studies have not found this to be a useful predictor (Mirak & Rescorla, 1998; Paul, 1996; Roulstone, Peters, Glogowska, & Enderby, 2003). Oro-motor control has been strongly associated with concurrent expressive language in toddlers (Alcock & Krawczyk, 2010), and has also had some success as a predictor of later language outcomes (Cleary, 2002).

In contrast, a concomitant early delay in receptive language has been found to be a moderate predictor of language outcomes in a range of studies, and is one of the stronger predictors identified to date (Chiat & Roy, 2008; Ellis Weismer, 2007; Henrichs et al., 2011; La Paro et al., 2004; Lyytinen et al., 2001; Liane Smith, 1998; Thal, Tobias, & Morrison, 1991; Zambrana et al., 2013). This is presumably because a delay in receptive language indicates the child has a more general weakness in their linguistic system, rather than potentially only having difficulties with speech production. However, the predictive accuracy of receptive language does not always hold on an individual level. For example, Ellis Weismer, Murray-Branch, and Miller (1994) reported their late talker with the lowest receptive language at intake had the best outcomes, and Paul, Looney, and Dahm (1991) reported that early receptive delays did not necessarily result in adverse language outcomes.

Early non-verbal IQ has had mixed success as a predictor of later language outcomes. Paul (1991) reported that while the late talkers in her cohort scored lower in non-verbal IQ than her TD controls, she concluded this was more likely due to the advantage that better language gave to performance on the Scales of Infant Mental Development (Bayley, 1969), rather than a genuine difference in non-verbal abilities between the two groups. A major twin study indicated there seems to be a general independence of genetic effects on language and cognitive abilities at age two years, despite some interrelationships (Price et al., 2000). Some studies have reported that early non-verbal IQ was not predictive of later language outcomes in their cohorts (Rescorla & Roberts, 1997; Rice et al., 2008) whereas others reported that it did add predictive accuracy (Ellis Weismer et al., 1994).

Several studies have found early gestural skills to be correlated with language outcomes (Rowe et al., 2012; Thal et al., 1991). However a large scale community sample by Zambrana et al. (2013) found that while imitative action predicted unique variance in expressive language outcomes between 18-36 months of age, gesture use (pointing) was not a significant predictor. Early symbolic play has also been significantly correlated with expressive language outcomes in some studies (Lyytinen et al., 2001; Maatta et al., 2012; Thal et al., 1991) but not to the extent of being a definitive predictor.

Late talkers have been found to have lower social competence and responsiveness compared with their TD peers (Bonifacio et al., 2007; MacRoy-Higgins & Kaufman, 2012; Paul & Shiffer, 1991), however this was not found to be predictive of language outcomes in Paul’s cohort (Paul, 1996). In contrast, Rescorla and Merrin (1998) found that their late talkers with a stronger intent to communicate had worse outcomes over a period of one year. The reasons for this were not clear, but the authors suggested that these children had a more severe underlying dysfunction, which meant they struggled to acquire expressive language despite having a strong desire to communicate.

While there are more developmental factors which have been investigated as predictors of language outcomes from toddlerhood, two conclusions can be drawn from the synopsis above. Firstly, when multiple studies have been done on a single predictor, a pattern of equivocal findings emerges. Secondly, even the best developmental predictors to date (such as gesture use, symbolic play, social skills and receptive language) have only a moderate success rate at predicting outcomes at a group level and less success at predicting individual outcomes. There are many methodological differences among these studies which could account for some of the discrepancies in findings, such as different measures used for the same construct; different ages at initial and outcome assessments; different sample sizes and composition; different criteria for delay and impairment and differences in statistical analyses. Another possible reason for these inconsistencies is that these developmental skills are results downstream of a variety of foundational inputs and learning processes, all of which can vary and interact at a causal level. Therefore environmental and genetic factors thought to influence variation in language ability also need to be considered as predictors.

1.1.3 Environmental factors as predictors of language outcomes

While genetic factors are critical in the phenotypic outcome of disorders, environmental stimulation also plays a powerful role in that children learn language from social interaction. Research into the effect of environmental stimulation on language outcomes has focused on two factors: the conversational styles of parent-child interactions and the quality and quantity of linguistic input.

Several studies have examined whether the interactions between late talkers and their parents are different in a way that may have negatively affected their language development. However the interaction styles of parents of TD children and parents of late talkers appear to be more similar than different (Paul & Elwood, 1991; Vigil, Hodges, & Klee, 2005). For example, Rescorla, Bascome, Lampard, and Feeny (2001) found that the only differences in conversational patterns between mothers of late talkers and mothers of TD children were that mothers of late talkers produced significantly more utterances and asked more questions. The authors concluded it was more likely that the differences noted arose because mothers of late talkers adjusted their conversational style to the level of their child's language, rather than that these differences in interaction style caused the delays.

Yet convergent streams of research provide evidence that the quality and quantity of linguistic input in the early years has an effect on children's vocabulary development. Hart and Risley (1995), in their large study of TD children, emphasised the importance of the cumulative effect of linguistic input over time, with the children with more enriched input not only having larger vocabularies at every time point, but also faster growth. The main factor associated with the level of linguistic input in this study was SES. Families with lower SES provided less linguistic input, which was also of lower quality. Other factors tending to coexist with low SES were lower maternal education, increased maternal stress, poorer maternal mental health, chaos / instability in the home and larger family sizes. All of these factors are thought to affect the amount of stimulating linguistic experiences very young children experience at home and have been implicated (on a group level) in language outcomes in three to eight year olds (Horwitz et al., 2003; La Paro et al., 2004; Reilly et al., 2010; Stanton-Chapman, Chapman, Bainbridge, & Scott, 2002; Taylor, Christensen, Lawrence, Mitrou, & Zubrick, 2013; Vernon-Feagans, Garrett-Peters, Willoughby, & Mills-Koonce, 2011).

However, studies have only found small associations (at most) between these environmental factors and late talking. Schjølberg, Eadie, Zachrisson, Oyen, and Prior (2011) found a small effect of similar family variables on expressive vocabulary at 18 months (the odds ratios for significant variables ranged from 1.09-1.26), despite their very large sample size (over 42,000). Two recent large community samples found SES and maternal education were not associated with late talking at two years of age (Reilly et al., 2007; Zubrick et al., 2007). Both of these studies reported higher vocabularies for first-born children; however birth order was not a significant predictor of late talker status in either cohort. These findings suggest that there are strong biological factors influencing the development of early expressive vocabulary. In support of this conclusion, Dale et al. (1998), in a large twin study, reported that genetic factors accounted for 73% of the variance in vocabulary development of two-year-old children in the lowest 5% of language ability, with 27% being accounted for by environmental effects. However for the entire sample of children, the balance was in the opposite direction, with 25% of the variance in vocabulary being accounted for by genetic factors. This evidence indicates that late talking is primarily genetically determined. These findings are supported by Zubrick et al.'s (2007) finding of a family history of late talking and Reilly et al.'s (2007) findings that family history of speech and language difficulties were significant predictors of low expressive language status at age 24 months.

In terms of an interaction, SES status seems to modify the heritability of general verbal and cognitive ability over time. A study of seven-year-old twins indicated that for twins from low SES families, 60% of the variance in verbal and non-verbal IQ was attributable to the shared environment, and virtually none was attributable to genetic factors, whereas the opposite pattern was found for twins from high SES families (Turkheimer, Haley, Waldron, D'Onofrio, & Gottesman, 2003). These findings suggest that the difference between impoverished and adequate environments is greater than the difference between adequate and enriched environments. Finally, it is possible that children who are genetically endowed with excellent language learning potential might display resilience in the face of mild environmental disadvantage, but those with a poor language learning ability may be more adversely affected by this disadvantage over time (Reilly et al., 2010).

1.1.4 Genetic factors as predictors of language outcomes

The search for specific genes causing neurodevelopmental disorders has proven more complex than initially anticipated. The concept of "a gene for language" is a popular but overly simplistic notion. Early research into genetic causes of LI focused on the KE family, who have a high familial incidence of Childhood Apraxia of Speech (CAS). While CAS is a broader phenotype than LI, covering speech-motor deficits as well as language difficulties, this example serves to illustrate the two main genetic ways that neurodevelopmental disorders can arise. Evidence showed the FOXP2 gene was deficient for all affected KE family members (Fisher, Vargha-Khadem, Watkins, Monaco, & Pembrey, 1998). However as research progressed, mutations of FOXP2 were found to be a fairly rare type of case of CAS, accounting for less than 2% of cases. It was found that monogenetic mutations such as this one, where a disorder can be accounted for by a variant of a single gene, typically result in a severe "all or none" disorder. Furthermore, Fisher reported that FOXP2 is rarely responsible for LIs in the more general population. More commonly, it is thought that the combined effects of several subtle variations in genetic codes across different chromosomal locations elevate an individual's risk of developing a disorder. These variations give rise to differences in the neural circuitry in individuals which are relevant for the development of a LI (Fisher, 2007). Work to identify a range of genetic differences which act as risk factors is being undertaken, not only for SLI, but also for related disorders such as autism, speech sound disorders, ADHD and dyslexia. Studies have found associations between the CNTNAP2 gene and late talking in two year olds, autism and poor non-word repetition skills (a heritable behavioural marker of SLI) (Vernes et al., 2008; Whitehouse, Bishop, Ang, Pennell, & Fisher, 2011). It is possible differences in this gene may contribute to the development of late talking and LIs.

While this is a promising avenue of research, genetic testing is not yet widely available to clinicians. Clinical research however can focus on abilities downstream of genetic effects, such as non-word repetition, which has been identified as a phenotypic marker of SLI (Bishop, North, & Donlan, 1996). Non-word repetition is thought to measure primarily phonological short term memory (PSTM) and will be considered as a predictor later in this review. Bishop, Holt, Line, McDonald, McDonald and Watt (2012) reported that mothers' non-word repetition scores when their children were aged 20 months improved prediction of late talkers' language outcomes at age four years. This sample was loaded with children with a positive family history for language or literacy problems and this may have influenced these results. A family history of speech, language and or learning difficulties has been noted as a risk factor for both late talking and LI (Choudhury & Benasich, 2003; Reilly et al., 2007; Zubrick et al., 2007). The final genetically determined factor to be noted is sex. However, this is of little use in predicting language outcomes, as while boys have a higher incidence of late talking at age two years, they are only marginally more susceptible for the development of LI over time than late talking girls (Rice et al., 2008).

1.1.5 Multiple risk factor models

There have been many review papers published which aimed to provide a predictive model of ongoing language delays based on prior research of late talkers. The first such article was that of Olswang, Rodriguez and Timler (1998). Only the four most recent reviews will be included here as a summary of the best predictive models available to date (Desmarais, Sylvestre, Meyer, Bairati, & Rouleau, 2008; Ellis & Thal, 2008; Paul & Roth, 2011; Rescorla, 2011). All four studies recommended a multiple risk factor model. The main risk factors identified by these reviews are listed in Table 1.1 below. This table is organised by risk factors and model, so a comparison of which factors are included in each model can be made. The five factors included in each model are delayed symbolic play, family history or speech and or learning difficulties, reduced language stimulation, poor social skills and limited gesture use. Desmarais et al. (2008) did not include receptive language as a risk factor, as the evidence they reviewed at the time on receptive delays was only based on approximately 50 late talkers. Since then, there have been larger studies which have confirmed the importance of receptive language as a predictor (Chiat & Roy, 2008; Henrichs et al., 2011; Zambrana et al., 2013).

Table 1.1

Comparison of the latest four multiple risk factor models for ongoing delays in late talkers

Risk Factors	Desmarais et al. (2008)	Ellis and Thal (2008)	Rescorla (2011)	Paul and Roth (2011)
Delayed symbolic play	√	√	√	√
Family history of speech and / or learning difficulties	√	√	√	√
Reduced language stimulation in the home (and associated factors)	√	√	√	√
Poor social skills / limited peer interaction	√	√	√	√
Limited or no gesture use	√	√	√	√
Otitis Media	√	√		
Delayed receptive language		√	√	√

Risk Factors	Desmarais et al. (2008)	Ellis and Thal (2008)	Rescorla (2011)	Paul and Roth (2011)
Reduced communicative intent / low rate of commenting	√			√
Limited phonology or vocalisations	√		√	√
Limited response to name and language				√
Severity of initial expressive delay			√	
Slow growth in expressive vocabulary	√			
Poor fast mapping	√			
Behaviour problems	√			
Parent needs		√		
Few spontaneous imitations				√
Lexical processing speed			√	

These studies report the most likely combination of risk factors to aid prediction of language outcomes in late talkers, based on their reviews of the literature. However these models have not actually been tested to determine their predictive accuracy. Group level associations may not translate well into predicting individual outcomes on a dichotomous level (e.g. LI / TD). In the absence of reports of the diagnostic accuracy of these models for predicting language outcomes, it is unclear what confidence a clinician could place in them. Only two late talker studies which predict language outcomes at age four years from variables measured at age two years using logistic regression have been published (Dale et al., 2003; Reilly et al., 2010). These studies used different measures, criteria for group status and reported their regression results differently from each other. While a direct comparison cannot be made, these brief summaries of each study illustrate the challenge of predicting dichotomous outcomes for late talkers.

Firstly, Dale et al. (2003) used logistic regression to predict the language status of four-year-old twin children (LI or TD) from their two-year-old assessment data (*n* = 8386). Although the relationships between two-year-old and outcome data were significant, the effect sizes were small. The ability of the regression equation to classify children as disordered or not was low. Their best model for predicting language outcomes included two-year-old expressive vocabulary, displaced reference, non-verbal cognition, sex and mother's education. This yielded a sensitivity of 51.5% and a specificity of 80%. In other words this model misclassified nearly half the children who would have LI and a substantial minority of children who were predicted to catch up, but did not (20%).

The other such study, Reilly et al. (2010) reported on the contributions of child, family and environmental predictors at age two years to language ability at age four years with a large community sample (*n* = 1596). Their study focused on measures of social disadvantage and demographics (child's age; sex; prematurity; birth weight and order; multiple birth; SES; non-English speaking background; family history of speech and or language difficulties; and maternal mental health, vocabulary, education and age at child's birth) as risk factors. Their logistic regression model using combined predictors was moderately successful in predicting those with low language at age four years (area under the curve = 0.76). By adding in late talker status at age two years, the predictions improved (area under the curve = 0.83).

Therefore as it stands, the best predictive models published are either unverified as a whole, or have low-to-moderate success rates at predicting language outcomes for two year olds. In order to improve this situation, additional factors involved in the development of LIs need to be considered. Moyle et al. (2011) argued that in light of the evidence from developmental, genetic and environmental factors included in previous studies, psycholinguistic processing variables were likely to be important in the development of LI. This line of reasoning will be summarised below, starting with theoretical accounts of SLI and then considering the research on the relationship between working memory and language development.

1.2 Theories of specific language impairment (SLI)

Theories of SLI can be divided into two main camps: domain specific and domain general. Domain specific accounts include a specific grammatical deficit (van der Lely, 2005) and delayed maturation of a constraint on grammatical computation (Rice, Wexler, & Cleave, 1995). These theories are challenged in their pure form by the existence of non-linguistic impairments in individuals with SLI. While the term ‘specific’ is part of the definition of SLI, non-linguistic skills are also known to be impaired. For example, children with SLI have been found to have depressed non-verbal IQ scores (Johnston, 1994); deficits in hypothesis testing (Kamhi, Catts, Koenig, & Lewis, 1984); cross-modal processing (Montgomery, 1993); voice processing abilities (Creusere, Alt, & Plante, 2004); sustained attention (Finneran, Francis, & Leonard, 2009); processing capacity and speed (Leonard et al., 2007); working memory (Montgomery, Magimairaj, & Finney, 2010) and motor coordination (Hill, 2001). The idea that a neurodevelopmental disorder could manifest itself in a single specific cognitive domain is not supported by recent learning regarding the nature of neurocognitive development in children (Karmiloff-Smith, 2009). The domain specific theories also fail to account for the range of linguistic difficulties these children have in addition to their grammatical errors. For these reasons, I will focus on domain general theories of SLI.

Domain general theories implicate poor temporal auditory processing (Bishop, Bishop, et al., 1999; Bishop, Carlyon, Deeks, & Bishop, 1999; Oram, 2003; Tallal & Stark, 1981), reduced working memory capacity, either specific to phonological processing (Chiat, 2001; Gathercole & Baddeley, 1990) or more broadly for linguistic processing (Baird, Dworzynski, Slonims, & Simonoff, 2010; Ellis Weismer & Evans, 2002; Montgomery & Evans, 2009), a generalised slow speed of processing (Dodd & Crosbie, 2002; Ellis Weismer & Hesketh, 1996) and weaknesses in the procedural memory system (Ullman & Pierpont, 2005).

It is beyond the scope of this study to consider all four causal theories in depth. In any case, they may be better viewed as being aligned than as competitors. All four theories are capacity limitation theories of LI; that is, a deficit in lower level processing is thought to constrain language acquisition. A commonality in three of these theories (reduced working memory, slowed processing speed and the procedural deficit hypothesis) is deficits in working memory or the prefrontal cortex. Domain general theories emphasise the cascading effect of lower level processing skills on the development of higher level abilities such as language. A pattern of weaknesses across the working memory system in two year olds could conceivably be a powerful predictor of both concurrent and later language status. Therefore I will focus on the role of working memory in language acquisition, exploring both its possible role in late talking and its utility as a predictor of toddlers' language outcomes over time.

Chapter Two: Working memory and language acquisition

2.1 Introduction to working memory

2.2 Baddeley and Hitch’s model of working memory

2.3 Phonological short term memory

2.3.1 The role of phonological short term memory (PSTM) in language acquisition

2.4 Declarative-procedural memory systems

2.5 Processing speed

2.6 Central executive (CE)

2.6.1 Verbal working memory (VWM)

2.6.1.1 The role of verbal working memory (VWM) in language processing

2.6.1.2 Evidence linking verbal working memory with language acquisition

2.6.2 Visual-spatial working memory (VSWM)

2.6.3 Executive functions (EF)

2.6.3.1 The relationship between executive functions and language acquisition

2.6.3.2 Inhibition

2.6.3.3 Shifting

2.6.3.4 Metacognitive executive functions

2.6.3.5 Plan / Organise

2.6.3.6 Emotional Control

2.7 Summary of literature review

2.1 Introduction to working memory

There are a multitude of papers on working memory utilising many different concepts and models. I will focus on the role of working memory in language acquisition, exploring its possible role in late talking and in predicting toddlers' language outcomes over time. Firstly I will introduce several constructs which will be frequently referred to throughout this study. The concept of new information being stored temporarily before being encoded to long term memory (LTM) is well accepted. There are two constructs which have short term storage as components: short term memory (STM) and working memory. STM refers to the temporary storage of items (e.g. images, sounds etc.) for a brief period of time (seconds). Items stored in STM are subject to rapid forgetting; most authors interpreting this decay as a function of time (e.g. Baddeley (2007) and Cowan (2008)), but some arguing for cue driven retention (Nairne, 2002). In this review, "working memory" is used to refer to the direction of attention for processing of information stored in short and long term memory, although some authors use this term to refer to storage alone.

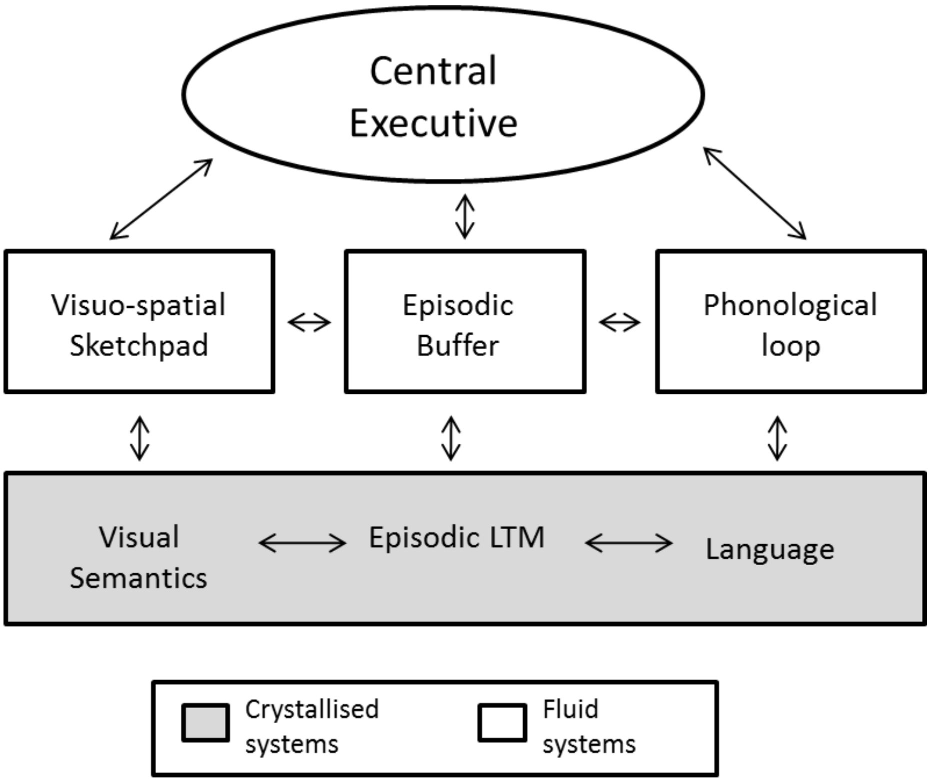
A discussion of how working memory models can shed light on language processing and acquisition follows. The most widely known working memory model is that of Baddeley (2007), which has been adopted as the main model for this research. Other working memory models differ in their scope and focus and in the number of dimensions in which individual variation in working memory can arise (Cowan, Rouder, Blume, & Saults, 2012; Ericsson & Kintsch, 1995; Just & Carpenter, 1992; MacDonald & Christiansen, 2002; Oberauer, 2009; Waters & Caplan, 1996). Much of the debate around models of working memory can be attributed to a difference in emphasis, terminology and scope of research rather than incompatibility of ideas (Baddeley, 2012). It is beyond the scope of this study to detail all the models of working memory or to try to prove the superiority of one over another. Instead, using Baddeley's model as a reference point, I will consider the contributions of several models to our understanding of how working memory and language interact.

2.2 Baddeley and Hitch's model of Working Memory

Baddeley and Hitch's model was originally published in 1974. The most recent thorough account of its development was presented in Baddeley (2007). It is a multi-store model, made up of a central executive (CE) command system and two slave subsystems; the phonological loop and the visuo-spatial sketchpad. The phonological loop stores verbal representations; the visuo-spatial sketchpad is its counterpart for visual and spatial information. Both of these subsystems are STM systems. The functions of the phonological loop are rehearsal and passive storage. Processing is directed by the CE, which is proposed to consist of four main executive functions (EF) of attention: dividing, switching and focusing attention and the ability to access LTM for working memory. These functions are thought to work in alliance to access and process information stored in the phonological loop, visuo-spatial sketchpad and LTM together to produce an appropriate response to stimuli. This entire process is called working memory. Another passive storage area called the "episodic buffer" allows for the integration of cross model information from both short term stores and LTM. Baddeley proposed that the episodic buffer is the basis of conscious awareness or thought.

While appearing to be highly modular in terms of its function, the bidirectional arrows between modules represent a high degree of integration and influence between components. In this way the model covers broad ranging cognitive functions with simplicity and flexibility, which probably accounts for its long standing popularity. For example, Baddeley sees the buffers as being highly influenced by LTM (for example, previously learned language), so incoming information is processed in a way that is influenced by prior learning. In this way, working memory is an interface. It processes information from a range of modalities and states. It also influences both coding to LTM and responses to tasks and events (Baddeley, 2012).

Figure 2.1 Baddeley's model of working memory[1]



Baddeley's model is based on evidence from neuroimaging and experimental studies with adults with brain lesions and children from a range of populations and factor analytic studies (Baddeley, 2007). Experimental evidence of different components has been gained by demonstrating that dual task demands do not disrupt the cognitive processing of each component. That is, if the addition of a second task does not significantly impact on the performance of the first, it is assumed to be using a different cognitive capacity. For example, random number generation is thought to tap the CE, while repeating a word is thought to use the phonological loop alone. In addition, evidence for double dissociation for all components has been reported, such as patients with intact CE functioning but poor PSTM, and vice

[1] From "Working Memory: Theories, models and controversies" by A. Baddeley, 2012, Annual Review of Psychology, 63(1), p. 16. Copyright (2012) by Annual Reviews. Reprinted with permission.

versa. The exception is the more recently added episodic buffer, which by its nature is difficult to isolate. See Baddeley (2007) for details of the evidence for this model.

Baddeley's model allows for the possibility that language or aspects of working memory can be impaired independently of the other and that poor working memory capacity constrains language acquisition and vice versa. Baddeley's model was also chosen as the majority of studies on paediatric working memory and language have used this model and retaining the same framework allows for easier comparisons across studies.

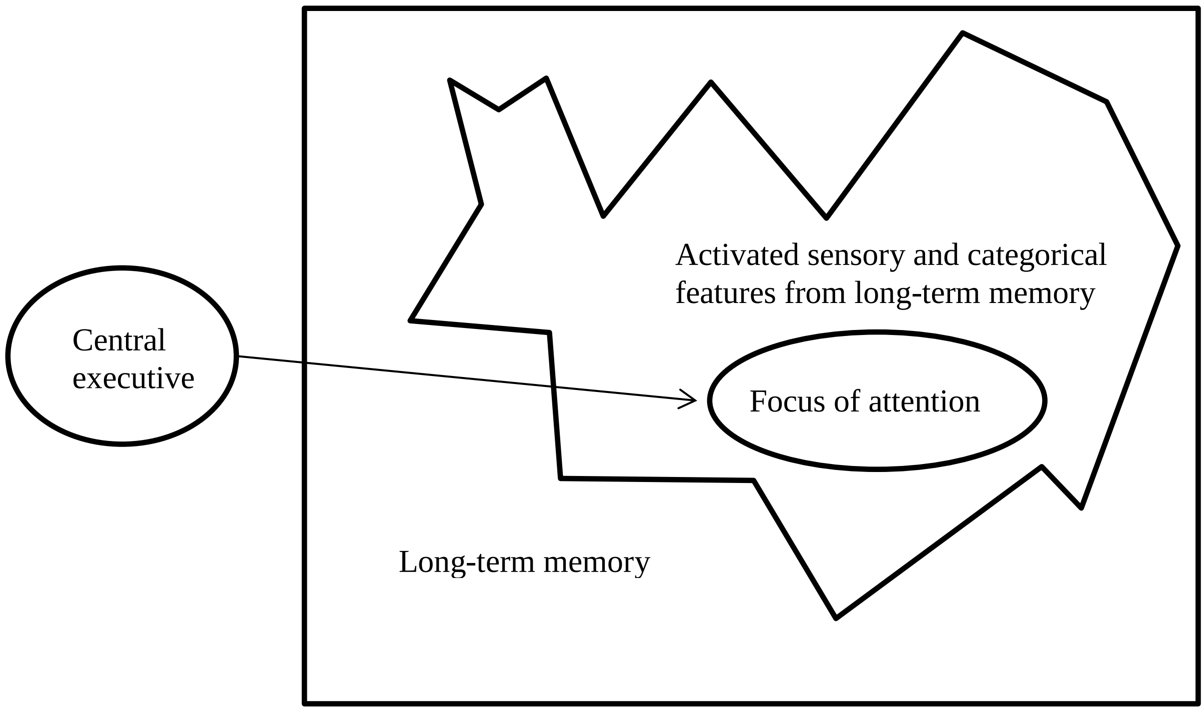
The main components of Baddeley's model which contribute to language learning are the phonological loop, CE and previously learned language stored in LTM. These components of the working memory system (as well as processing speed) have been investigated with regards to their role in language processing in a variety of populations. The following section briefly summarises what is known about each component's role in language processing and then considers the role it may play in early language acquisition. Additional perspectives from other models of working memory are discussed throughout where informative.

2.3 Phonological short term memory

Phonological short term memory (PSTM) (the capacity of the phonological loop) is typically measured using span tasks or tests of non-word repetition. Span tasks, such as word recall or digit recall, involve repeating back strings of words or digits which increase in length. Non-word repetition tasks have been devised which increase by syllable length from one to five syllables and which vary in terms of phonological complexity and word-likeness. Gathercole (2006) proposed that non-word repetition is useful as an index of the quality of phonological storage. She acknowledged that this quality is determined by several factors, namely the familiarity of the segments, how quickly the representations fade in an individual, the length of the stimulus and phonological similarity. Each of these factors is expected to influence the learning of new words. Non-word repetition performance relies on intact auditory processing, phonological analysis, phonological storage and speech motor planning and output processes (Gathercole, 2006).

In order to understand the idea of poor STM capacity constraining language processing, it is necessary to consider the links between STM and LTM in more detail. Cowan's work has focused on the nature of the interaction between the CE and long and short term memory (Cowan, 2008).

Figure 2.2 Cowan's model of working memory[2]



"Working memory" in this model refers to the amount of information in LTM held in an accessible state for a short period of time (STM), as directed by the CE (shown in Figure 2.2). Cowan relates this to a pattern of neuron firing or cell assembly which represents a particular idea. As long as the neurons are active, the idea is held in STM. The focus of

[2] From "Evolving concepts of memory storage, selective attention and their mutual constraints within the human information processing system" by N. Cowan, 1988, Psychological Bulletin, 104, p. 180. Copyright (1988) by the American Psychological Association. Adapted with permission.

attention is the section of representations where the most mental energy is directed. The activated portion of memory is subject to decay and interference. Individual differences arise in the amount of activation a person can maintain and also the effectiveness of executive control of attentional resources. In Cowan's model, working memory is equivalent to STM when it is being used to solve a problem or to do a task. Cowan proposed that information held in STM becomes a LTM trace, if the pattern of neuronal firing becomes strong enough, that it can be reactivated over longer periods of delay.

The relevance of this account to the current study is that it implies there will be an impact of STM capacity on coding to LTM and vice versa. If a person has a limited PSTM, it will take longer to code novel words to LTM, as the individual will not be able to maintain as many neurons in an active state at any one time. In addition, it also suggests that LTM will be a support for STM storage through a process called "chunking" (Cowan, 2008). This is demonstrated through the word-likeness effect on non-word repetition tasks. Individuals can repeat non-words which are similar to real words with greater accuracy than non-words which are dissimilar (e.g. "trumpetine" is easier to repeat than "skiticult") (Gathercole, 1995). This indicates that the phonological system in LTM supports the function of the phonological loop. The more word-like a non-word is (the more chunks of the word that are familiar), the more a child can use their LTM store of vocabulary to retain the form of the word (Gathercole, 1995). In this way the role of increased experience with language can be seen to improve working memory capacity. This has implications for the interpretation of associations between measures of non-word repetition and language development, which will be discussed throughout this thesis.

2.3.1 The role of phonological short term memory (PSTM) in language acquisition

The role of the phonological loop in language acquisition has been thoroughly investigated across a range of ages and populations. To summarise the literature, there is evidence to support the theory that the phonological loop plays an important role in establishing memory traces for new phonological forms, which through a process of refreshing or rehearsal, become a part of LTM (Baddeley, Gathercole, & Papagno, 1998). Adults with specific lesions resulting in limited PSTM have difficulty acquiring new phonological forms (such as foreign vocabulary), and with comprehension of syntactically complex and long sentences which are dependent on recall of serial order, while other aspects of their language learning system remain intact (e.g. associative learning) (Baddeley, Papagno, & Vallar, 1988). Moderate correlations have been demonstrated between PSTM and vocabulary in TD children aged four to eight years (Bowey, 2001; Gathercole, Willis, Emslie, & Baddeley, 1992); although several authors have warned against interpreting these data as evidence for a causal relationship between PSTM and vocabulary acquisition (Bowey, 2001; Melby-Lervag et al., 2012; Metsala, 1999). The phonological loop seems to also play a role in acquiring the phonological forms of morphosyntax (Adams & Gathercole, 1995; Blake, Austin, Cannon, Lisus, & Vaughan, 1994; Chiat & Roy, 2008). Poor PSTM skills have been consistently linked with SLI even compared with age- and language-matched controls (Archibald & Gathercole, 2007; Baird et al., 2010; Dollaghan & Campbell, 1998; Gillam, Cowan, & Day, 1995; Marton & Schwartz, 2003; Montgomery & Evans, 2009). Within the SLI population, few children have PSTM abilities within the normal range; however, when it does occur, such children typically have higher language and literacy performance than those with poorer PSTM (Alloway & Gathercole, 2006; Botting & Conti-Ramsden, 2001; Gathercole & Baddeley, 1990). Evidence was found in a large twin study for a strong genetic component to poor performance on a non-word repetition task. As previously mentioned, non-word repetition was suggested as a phenotypic marker for LI (Bishop et al., 1996).

The predominant interpretation of these data is that capacity limits in PSTM slow language acquisition. The main argument against this “capacity limits” interpretation of the research is that working memory skills arise alongside or as a result of language development, rather than as its cause. Metsala (1999) reported that the association between vocabulary and non-word repetition in her preschool sample was accounted for by a third variable, phonological awareness (these findings have been replicated by Bowey (2001)). This interpretation is in accordance with the ‘lexical restructuring hypothesis’, which proposes that as vocabulary increases, children’s word recognition skills move from holistic lexical representations to representations on a segmental level (e.g. syllables and phonemes). This in turn allows for the development of phonological awareness. Metsala argued that it was this process which facilitated more successful non-word repetition. Metsala however did not rule out the idea that the rate of vocabulary growth might be affected by PSTM.

The relationship between working memory and language is most likely bidirectional (Gupta & Tisdale, 2009), with the strength of influence changing over time. As PSTM is by definition, a phonological memory system, it should not be seen as separate to language. However evidence shows it can be stronger or weaker relative to other linguistic skills. Children with resolved early language delays have been found to continue to have poor non-word repetition in comparison to children without a history of such delays, both at a preschool and primary school ages (Bishop et al., 1996; D'Odorico, Assanelli, Franco, & Jacob, 2007; Thal et al., 2005). However this finding is not universal (Petruccelli, Bavin, & Bretherton, 2012). Children with SLI in the primary school years perform worse than younger language-matched controls at non-word repetition tasks as a group (Gathercole & Baddeley, 1990). These findings would not be possible if PSTM developed solely in response to language acquisition. Instead, the evidence suggests that as vocabulary increases, reliance on PSTM to acquire new words decreases, due to the increased support from LTM stores for analysis and storage of new incoming phonological information (Gathercole, 1995).

2.4 Declarative-procedural memory systems

The hypothesised role of PSTM in language acquisition needs to be seen in the context of broader learning systems, as multiple cognitive processes work in concert to allow language mastery. As outlined in Moyle et al. (2011), Ullman’s (2001) declarative-procedural model proposed two memory systems which are conceptually separable, but work together in language learning. The procedural system is responsible for the learning and storage of rule-governed knowledge for directing sequences of actions (e.g. motor-speech production and making gestures) and language behaviours (e.g. syntax and phonology). These types of linguistic information are sequential and probabilistic in nature, and are acquired by detecting regularities in input over time. This learning is called procedural or statistical learning. It is considered to be an implicit learning system as the process of learning is not available to conscious awareness. The procedural memory system is based in the frontal basal ganglia circuits, parietal cortex, superior temporal cortex and the cerebellum. The declarative memory system is responsible for learning chunks of knowledge of general relations about objects and their properties, such as form-referent mappings in vocabulary acquisition. It has been located in the medial temporal lobe network, particularly the hippocampus. Declarative learning is thought to be an explicit learning system. Both of these systems contribute to LTM. Ullman and Pierpont (2005) describe further details of these systems and their neural correlates.

Studies investigating statistical learning in early vocabulary acquisition have suggested that late talkers may have deficits in this area, and that this may be linked with poor PSTM. Statistical learning is a powerful learning mechanism in early vocabulary acquisition. Word boundaries are not marked by pauses in continuous speech (e.g. “herecomesyourbottle”). In order to identify word boundaries, infants must use statistical (probabilistic) cues in the input. These cues include sequential probabilities of syllable and word level combinations (Saffran & Wilson, 2003). Stokes (2010) demonstrated that toddlers’ first spoken words are influenced by the statistical regularities of the phonological and lexical characteristics of words in the ambient language. For example, toddlers’ first words tend to be short words that have high neighbourhood density. An example of a high neighbourhood density word is *cat* which has 35 phonological neighbours in British English, such as *mat, pat, cap, kit.* An example of word with a low neighbourhood density is *mouth* which has six neighbours: *math, mouse, myth, moth, south, mouth (verb)*. Words with higher neighbourhood density are thought to be easier to recall, as representations of similar words in LTM support the mapping of the new forms (Aslin & Newport, 2009). A deficit in PSTM would presumably affect the development of the expressive lexicon more than the receptive lexicon, as better formed phonological representations are required for production rather than recognition of words (Munroe, Baker, McGregor, Docking, & Arculi, 2012). This was found to be the case in a recent study of toddlers’ vocabularies. Late talkers (those scoring below the 16th percentile, but having more than 20 words in their total vocabularies) had higher mean neighbourhood densities in their expressive vocabularies than their receptive vocabularies, whereas the average talkers did not show this bias. This study also reported that late talkers at (age 23-24 months) resembled younger language-matched peers (aged 17-18 months) in the neighbourhood density of their lexicons (Stokes, 2013). This evidence supports the view that late talkers have a prolonged period of using high neighbourhood density as a cue to word production, possibly due to poor PSTM (Stokes, Kern, & Dos Santos, 2012).

Late talkers have also been found to lack sensitivity to the statistical properties of their language in novel word learning paradigms. MacRoy-Higgins, Schwartz, Shafer, and Marton (2013) found that late talkers were worse at comprehension, production and sensitivity to mispronunciations in novel words than TD children who were matched by age, gender, SES and maternal education. Unlike the TD children, who showed better performance for words with high neighbourhood density, the late talkers performed similarly on words with both low and high neighbourhood density ratings. The authors concluded that this lack of sensitivity to the statistical properties of the novel words likely meant late talkers lacked detailed phonological representations of the novel words. This was despite the fact that the late talkers scored in the average range for receptive vocabulary. Therefore this study also points to a possible deficit in PSTM in late talkers.

PSTM has also been linked to variability in toddlers' concurrent expressive language skills and later language outcomes. Stokes and Klee (2009b) found that non-word repetition accounted for 36% of the variance in concurrent expressive vocabulary in their sample of 232 toddlers. Only toddlers who completed the test of non-word repetition were included in the analysis (77%). Hoff, Core, and Bridges (2008) assessed a total of 36 children aged between 20 and 24 months of age. They reported non-word repetition predicted vocabulary over and above real word repetition, indicating that the non-word repetition task measured something in addition to word articulation ability, purportedly memory skills. Chiat and Roy (2008) investigated the predictive validity of word repetition tasks in their clinical sample of 163 2;6-3;6 year olds. These children's morphosyntax outcomes at four to five years old were best predicted by the early real and non-word repetition measures. While this association between early word repetition and later morphosyntax was no longer seen when this cohort was reassessed at ages nine to eleven years (Chiat & Roy, 2013), the authors noted that all the children with deficits in morphosyntax at 9-11 years had had severe difficulties with word repetition at intake. However, by this age, it seems either some children had been able to compensate for this difficulty, or the high level of phonological difficulties in the cohort initially may have meant early word repetition deficits were over identified. However, overall these studies demonstrate a strong association between non-word repetition and vocabulary in the toddler years, and indicate that PSTM could be a useful predictor of later language outcomes. It may be that late talkers with better PSTM skills resolve their language delays over time, while those with very poor PSTM at age two years continue to struggle with language. I therefore proposed to measure PSTM using a measure of non-word repetition in the current research.

2.5 Processing speed

Speed of processing is the amount of information that can be processed in a given unit of time. It is measured from the time a stimulus is presented, to the time the individual's response is registered. The speeds of different aspects of cognitive processing (e.g. non-linguistic, motor and linguistic processing) can differ within individuals (Leonard et al., 2007). Processing speed increases as a function of chronological age until adulthood. It is thought to be a key determiner of cognitive development. C. Miller, Kail, Leonard, and Tomblin (2001) found that the association between age and other cognitive processes (such as executive functions (EF), attention and PSTM) was mediated by, or the same as, the relationship between age and speed of processing in TD children. As language is a rapid transient auditory code, it is assumed that a slower processing speed may affect the likelihood of linguistic input being encoded to LTM and thus impact its speed of acquisition (Bayliss, Jarrold, Baddeley, Gunn, & Leigh, 2005). Therefore, while processing speed is not explicitly included in Baddeley's model, it is an important component of the working memory system.

Slowed processing speed has been proposed as a causal theory of SLI as previously mentioned. Hayiou-Thomas, Bishop, and Plunkett (2004) simulated capacity limitations (such as those caused by slowed processing speed) in TD primary school aged children by increasing the rate of presentation of linguistic stimuli. These children then displayed linguistic errors similar to those of children with SLI. Primary and secondary school aged children with SLI have been reported to have slower processing times than TD children on a range of linguistic and non-linguistic tasks (Ellis Weismer & Hesketh, 1996; Leonard et al., 2007; C. Miller et al., 2001; C. Miller et al., 2006; Montgomery & Evans, 2009). However, some studies have found similar reaction times between children with SLI and various control groups (Marton, Campanelli, Scheuer, Yoon, & Eichorn, 2012; Noterdaeme, Amorosa, Mildenberger, Sitter, & Minow, 2001; Spaulding, 2010). It seems that slower processing is a feature of children with SLI as a group, but not every individual with SLI has slowed processing.

Processing speed has been linked with later language outcomes in both TD and late talking toddlers. Marchman and Fernald (2008) reported the speed of spoken word recognition at 25 months correlated with language abilities in the same cohort measured at age eight years. In addition, the same research team reported that faster processing speeds predicted language resolution in a cohort of late talking toddlers (Fernald & Marchman, 2012). This is therefore a promising predictive factor to include in this study. To extend Marchman and Fernald's work, I proposed to include a measure of processing speed in the current research alongside measures of working memory to see if it contributes unique variance in later language outcomes.

2.6 Central executive (CE)

Given that the CE is thought to direct attention in various ways across domains for a range of learning tasks, it is difficult both to define and study. The CE has been criticised as being the least well defined component of Baddeley's model (Miyake et al., 2000). In this study, its relationship with language acquisition will be considered from three viewpoints to ensure breadth of coverage. Firstly, the role of the CE in directing attention for verbal working memory (VWM) and visual-spatial working memory (VSWM), and then the relationship between broader EF and language acquisition (using Isquith, Gioia and Epsy's (2004) model of EF).

2.6.1 Verbal working memory (VWM)

In Baddeley's model, VWM involves the phonological loop as a passive store and the CE to direct attention between the phonological loop and LTM stores of language. Thus it is the active processing of both novel and previously learned verbal information. It is important to note that deficits in PSTM will also have an effect on VWM, but the primary focus here is the function of the CE. Listening span tasks are commonly used to assess VWM. They involve a two-step process: firstly a simple semantic judgement about a series of spoken sentences (e.g. true or false) and then recalling the final words of each sentence. An individual's listening span is the number of sentences they can accurately process and recall. There are alternative VWM tasks, such as the backwards span task, where participants repeat a string of digits back in reverse order (Alloway, 2012).

VWM has been less extensively researched with regards to its role in language acquisition and LI than PSTM, and the results are more difficult to interpret. Initial research using the listening span task found it predicted verbal IQ in college students better than PSTM measures (Daneman & Carpenter, 1980; Gaulin & Campbell, 1994). Similar findings have been reported in TD primary school students (Gathercole, Tiffany, Briscoe, Thorn, & team, 2005). In school-aged children, VWM has also been linked with metaphor interpretation and conjunction use (Johnson, Fabian, & Pascual-Leone, 1989), pragmatics (Freed, Lockton, & Adams, 2012), and comprehension of narratives (Montgomery, Polenenko, & Marienellie, 2009; Zaretsky, 2004). Of main interest here, evidence suggests a role for VWM in sentence comprehension, particularly complex sentences with unfamiliar or novel syntax for that individual, in both TD children (Magimairaj & Montgomery, 2012a, 2012b) and children with SLI (Montgomery & Evans, 2009).

Once again the complexity of the language learning system must be considered. The interaction and relative importance of VWM, PSTM, procedural and declarative memory systems for language acquisition is largely unknown. Oberauer (2009) described how the parsing and associative routines which operate as part of VWM to allow language comprehension could be considered to be a part of procedural memory. This model demonstrates how implicit and explicit learning systems may interact in language learning. That is because procedural learning is predominantly implicit and VWM processing is predominantly an explicit process. Such interactions are seldom mentioned in the VWM literature, but are fundamental to all learning (Sun, Zhang, Slusarz, & Mathews, 2007). In a recent review of the VWM literature, Kidd (2013) called for more carefully controlled experimental studies investigating this interaction of implicit and explicit learning processes. As a step in this direction, Boyle, Lindell, and Kidd (2013) investigated the role of the CE, episodic buffer and phonological loop in sentence comprehension in four- to six-year-old children. They found VWM predicted comprehension using regression modelling until the measure of the episodic buffer (a sentence repetition task) was added to the model. They then found that sentence repetition interacted with the canonicity of the sentence, and VWM was no longer a predictor. These findings emphasise the role of frequency of input in developing proficiency in learning different forms of language (canonical word order being more frequently occurring than non-canonical word order). The authors suggested that sentence repetition was a metric of the child's ability to learn language from statistical regularities in ambient language (accounting for the interaction with canonicity), and that this was a stronger predictor of sentence comprehension than VWM. The authors suggested further investigation of the validity of sentence repetition tasks as a measure of the episodic buffer was needed. Further questions could be raised regarding the validity of sentence repetition as a measure of statistical learning. Taken at face value however, this study lends support to the idea that while VWM is associated with language comprehension, it may not be the most powerful language learning mechanism involved. Statistical learning through implicit mechanisms (procedural memory) may play a greater role, but confirmation of this awaits a unified model of language learning and is beyond the scope of this study. However, I will explore the associations between VWM and language processing and relate these to early language acquisition.

2.6.1.1 The role of verbal working memory (VWM) in language processing

VWM may play a different role in language processing across different stages of development. Patients with Alzheimer's disease (who have very limited VWM spans) can converse easily (Waters, Caplan, & Rochon, 1995). This seems to challenge the view that VWM is a vital part of language processing. Several investigations of working memory have suggested that comprehension and use of new language forms may require conscious, effortful processing in the first instances, and gradually proceed to automatic cognitive routines as the individual proceeds to mastery (Ericsson & Kintsch, 1995; Waters & Caplan, 1996). As previously mentioned, Oberauer (2009) suggested that procedural memory routines in LTM play this role of automatic sentence parsing and interpretation. Caplan and Waters (2013) speculated that when this fails, such as when language is not sufficiently familiar to be understood automatically, the phonological loop might be used to replay the words in serial order. This would allow for conscious processing and manipulation of the sentence for possible interpretations by VWM.

To relate this back to the possible role of VWM in early language acquisition, VWM may play a role in the early stages of language development, as little of the language heard or used would be processed automatically yet. While VWM is not assumed here to be the most important mechanism for learning syntax, morphology or vocabulary, greater VWM capacity is hypothesised to give a developmental advantage in early language acquisition. Mainela Arnold, Misra, Miller, Poll, and Park (2012) argued against an explicit metalinguistic conscious learning process for early language learning, stating that these more abstract skills develop later once language is developed. It is not proposed here that a very young child would reason about which word ending to use, as an adult might in second language acquisition. However, a child's conscious analysis of an experience, even at a rudimentary level, could only aid their implicit mapping of the language used to the correct elements of the experience. The ability to hold more spoken language in mind while processing its meaning can only be an advantage in developing explicit and implicit comprehension skills. A longer VWM span at this age may facilitate better topic maintenance, leading to better conversations skills, resulting in more positive practise with language and faster progress to mastery. Finally, children who have poorer implicit learning mechanisms for language may use PSTM and VWM as compensatory mechanisms to support their development of language.

As for PSTM, gains in language can also be seen to drive improvements in VWM. MacDonald and Christiansen (2002) argued that VWM is an aspect of the linguistic system and cannot meaningfully be separated from it. According to this view, language acquisition is not constrained by a shorter VWM span, but rather VWM tasks measure the strength of the linguistic neural network itself. Instead individual differences in language arise from differing levels of experience with language and also differences in statistical learning abilities. In general support of this, Mainela Arnold et al. (2012) demonstrated that a phonological awareness measure of elision (segmentation) predicted listening spans over and above language, reading and non-verbal IQ abilities in primary school aged children. They interpreted this as Metsala (1999) did for PSTM, that increased linguistic ability enabled increased metalinguistic awareness of segmentation which resulted in higher scores in the listening span task. These authors argue that the idea that VWM contributes to language development is circular, as language development itself contributes to VWM. This argument has strengths in that by definition, VWM is part of the linguistic system and does use LTM to support its operations. However this does not preclude the possibility that aspects of the linguistic system (such as PSTM and VWM) have a function in language learning and that capacity limits in these memory systems could constrain learning in certain areas. Evidence of dissociation between VWM and language has already been mentioned with reference to patients with Alzheimer's disease. While this example shows an acquired processing deficit in a mature linguistic system rather than a developing one, it does indicate that VWM and wider language skills are dissociable. In addition, Newman, Malaia, Seo, and Cheng (2013) found evidence from their functional magnetic resonance imaging (fMRI) study of college aged students that during reading comprehension probes, there was neural activation in brain areas not specific to language, but instead related to executive processes. A summary of further evidence on the relationship between VWM and language in development is outlined below.

2.6.1.2 Evidence linking verbal working memory with language acquisition

Research with primary school aged children suggests a strong relationship between VWM and language at a group level. Florit, Roch, and Levorato (2011) found that VWM, vocabulary and verbal IQ all accounted for unique variance in listening comprehension scores in four- to six-year-old TD Italian children. Several studies have found primary school aged children with SLI perform more poorly on tasks of VWM compared with age- and language-matched controls (Archibald & Gathercole, 2006a; Henry, Messer, & Nash, 2012; Montgomery, 2000a, 2000b; Montgomery & Evans, 2009). Note that not all children with SLI scored low for VWM in these studies and that some TD children scored low as well.

These results suggest the view that working memory only increases as the linguistic network itself is strengthened is unlikely, as language and VWM show a level of dissociation in developmental populations.

Longitudinal studies using VWM as a predictor of language outcomes in children are rare. There has been some research in this area with other developmental populations. VWM skills in primary school aged children with cochlear implants and Fragile X were found to predict language outcomes which were measured one to several years later (Kronenberger et al., 2013; Pierpont, Richmond, Abbeduto, Kover, & Brown, 2011; Pisoni, Kronenberger, Roman, & Geers, 2011). There have been no studies using VWM as a longitudinal predictor for language in TD or late talking toddlers published to date.

The view that VWM may play a role in language acquisition has been challenged in two respects. Firstly the idea of a dual deficit in the CE and phonological loop in children with SLI was questioned by two studies which found that deficits in PSTM alone accounted for the low VWM scores (Briscoe & Rankin, 2009; Freed et al., 2012). However, these results could well have arisen from an interaction between the small sample sizes (n = 12-14) in these studies and the heterogeneity of the LI population. Secondly, Lum, Conti-Ramsden, Page, and Ullman (2012) did not find a correlation between VWM and vocabulary or grammar in either children with SLI or TD children, despite the fact the children with SLI had deficits in VWM over and above those expected by their language level. Instead they found a correlation between procedural learning and grammar in the TD group and declarative memory and grammar in the SLI group. The view that VWM may play a compensatory role in grammatical learning if procedural memory systems were poor was not supported by this study. Instead the declarative system seemed to compensate in the SLI group, as predicted by the procedural deficit hypothesis (Ullman & Pierpont, 2005). It is possible there were no correlations between grammar and VWM because of the children's stage of language development (mean age ten years). In support of this, there were no significant correlations between the phonological loop and vocabulary in this study either. These results do not preclude VWM playing a role when grammar is newly emerging (ages two to four years).

There is a dearth of research on the role of VWM in language acquisition in the preschool years. This is presumably because standard VWM assessments are too complex for many preschool children to participate in, particularly those with emergent language. Most children cannot participate in listening span or backwards digit span tasks until they are four to five years of age. Few alternative tasks for three to four year olds have been reported in the literature. Hughes (1998) used a “noisy book” task where the child was required to listen to a string of animal names, and then press the buttons which corresponded to these names. The buttons played the animal’s sound when pressed. This task was used with three to four year olds. Willoughby, Blair, Wirth, & Greenberg (2010) described a working memory task designed for three year olds based on recalling which animal(s) they had seen on a previous page. These two studies demonstrated moderate (but not always significant) correlations between VWM and language in the age range three to four years (Hughes, 1998; Willoughby et al., 2010).

There are currently no published tasks suitable for assessing VWM in two year olds, particularly those with limited expressive language. Evidence of the relationship between late talking and VWM is sparse and difficult to interpret. Two cohort studies have been used to explore this connection. Petruccelli et al. (2012) found in their study of five-year-old RLTs that these children did not have a lower VWM span compared with TD controls. However they commented that their measure of VWM (backwards digit span) may not have been an appropriate measure even at this age. In contrast, a follow-up study of RLTs at ages 13 and 17 years (Rescorla, 2009) reported significant group differences in working memory composite scores compared with a comparison group of children matched for age, SES and non-verbal ability. These composites comprised PSTM and VWM measures. This finding was reported at both ages tested. While the RLTs scored in the average range in working memory on all measures as a group, some individuals' scores were below the 10th percentile. In this cohort, Rescorla commented that tasks requiring processing and mental manipulation of verbal information were noted to be particularly difficult for the RLTs. However, due to the reporting of composite scores rather than individual working memory assessment results, it was not possible to distinguish between PSTM and VWM skills in this study. These results could therefore primarily reflect a deficit in PSTM.

This general pattern of results suggests that children with SLI in the primary school years show deficits in VWM beyond those expected by their lower language ability and that this could be interpreted as being causal, although the relationship is far from simple. To my knowledge, there have been no prospective longitudinal studies reporting the role of VWM in language acquisition of late talking or TD toddlers. VWM may play a role in early language acquisition and could be a useful predictor of language outcomes from toddlerhood. I propose to measure VWM in the current research using a novel task developed for this purpose.

2.6.2 Visual-spatial working memory (VSWM)

As the CE is usually considered to be a domain general resource (Baddeley, 2007; Cowan, 2008; Kane et al., 2004), deficits in the CE should also be detected in the visual-spatial modality. If the CE is involved in language acquisition, children with SLI should show deficits in both VSWM and VWM. Children with SLI have been shown to have intact visual-spatial STM skills, meaning any deficits in VSWM can be attributed to difficulties with the CE (Archibald & Gathercole, 2006b). Several well-designed studies of school aged children with SLI and various control groups have found this result (Henry et al., 2012; Marton, 2008; Marton et al., 2012). My literature search did not reveal any studies investigating VSWM in late talking toddlers. However, Petruccelli et al. (2012) tested RLTs at age five years on a range of memory tests and reported that RLTs did not have any deficits in VSWM skills.

The relationship between VSWM and language acquisition is not well understood. In the preschool years, studies using an early measure of VSWM (the A not B task) with TD children, children with developmental delay or autism in the age range two-to-five years have not shown a significant correlation with receptive vocabulary (Epsy, Kaufmann, McDiarmid, & Glisky, 1999; Epsy, Kaufmann, Glisky, & McDiarmid, 2001; Griffith, Pennington, Wehner, & Rogers, 1999). However, the studies led by Epsy had 30% and 19% missing data (respectively) on the receptive vocabulary measure due to fatigue effects. This could have skewed their results towards the higher end of ability and reduced the likelihood of finding a significant correlation. Therefore the association between VSWM and early language development remains an open question. Poor early VSWM, in so far as it implicates poor function of the CE, may function as a predictor of poorer outcomes in late talkers. Therefore a VSWM task was included in the current research.

2.6.3 Executive function (EF)

'Executive Function' is an umbrella term for the set of task-focused behaviours arising from the functioning of the CE. Attempts to define EFs have included the following constructs: shifting, inhibition, sustained attention, focusing attention, accessing LTM, updating working memory, dual tasking, emotional control and planning / organising. The results of Miyake et al.'s (2000) frequently cited factor analysis study of college aged students suggested that there are three basic EFs (inhibition, updating of working memory and shifting). Miyake et al. proposed that these three EFs work in concert to direct attentional resources, rather than in isolation. For example, shifting involves both working memory and inhibition. Therefore these EFs are best conceived of as distinguishable, but not entirely separable. A dual (working memory / inhibition) or unified model of EF for preschoolers has been suggested by some authors on the basis of their own factor analyses (Garon, Bryson, & Smith, 2008; Senn, Espy, & Kaufmann, 2004; Wiebe, Espy, & Charak, 2008; Willoughby et al., 2010). It has been questioned whether the finding of different models for children compared with adults is a methodological artefact of the tasks used to measure EF with children, rather than a genuine difference in the structure of brain function across different ages (M. R. Miller, Giesbrecht, Müller, McInerney, & Kerns, 2012). To further complicate this picture, results of a recent meta-analysis suggested that EFs may be best thought of as content-specific rather than function-specific. Nee et al. (2013) suggested that rather than EFs being distinguished by function (e.g. inhibition, updating, shifting etc.), two separate frontal lobe regions were activated according to whether the task contained verbal or spatial content. This alternative model for the CE will be considered further in the discussion.

These conflicting viewpoints make it difficult to know how to best measure EF, particularly in toddlers. Including multiple experimental measures of EFs would have made the test protocol proposed for the current research too taxing for toddlers. However, Gioia, Epsy, and Isquith (2003) developed a parent report questionnaire (the Behavioural Rating Inventory of Executive Functions – Preschool Version (BRIEF-P)) using a five-fold EF model (Isquith, Crawford, Espy, & Gioia, 2005; Isquith et al., 2004). Isquith et al.'s (2004) model covers the three fundamental EFs of inhibition, working memory and shifting, as well as the metacognitive functions of emotional control and planning / organising. This is the only EF parent questionnaire available which was designed for pre-schoolers and therefore was included in the test protocol for the current research. The disadvantage of this model is that it does not allow comparison of EFs by domain. However the inclusion of VWM and VSWM tasks already in the test protocol should allow for this comparison. Isquith et al's (2005) model will therefore be used as the basis of this review on the relationship between EF and language development.

2.6.3.1 The relationship between executive functions and language acquisition

Correlations between EF and language are commonly reported, and will be outlined in later sections for each EF in the BRIEF-P. These associations are typically interpreted as being unidirectional in causation as language skills are believed to enable better EF through improved use of language for self-talk and rehearsal. It has been suggested that rehearsal and self-talk strategies develop around age seven years rather than in the preschool years (Gathercole & Adams, 1993), meaning that this does not account for the relationship early on. However, at a preschool level, a greater ability to verbalise feelings and desires and to discuss behaviour and consequences with adults would logically help develop better EF. In any case, an equally valid interpretation of the observed correlations between language and EF is that EF supports language development. Four EFs from the BRIEF-P are reviewed in the following section (Inhibition, Shifting, Emotional Control and Plan / Organise) with regards to this hypothesis. The fifth part of this model ‘Working Memory’ has been discussed previously in the VWM and VSWM sections.

2.6.3.2 Inhibition

Inhibition is often referred to as a unitary construct in neurodevelopmental studies; however, it is better thought of as an umbrella term. While there is no consensus in the literature, there seems to be two main forms of inhibition: the ability to lower activation of mental schemes which are irrelevant or distracting to task performance, and the ability to delay or prevent a prepotent response (Marton et al., 2012). Resistance to interference can be measured in working memory tasks by analysing subjects’ error patterns, for example repeating stimuli from a previous item (Ellis Weismer, Evans, & Hesketh, 1999). Response inhibition is measured in a variety of ways, for example resisting the urge to take a treat before being allowed to; staying “as still as a statue”; or inhibiting a motor response which has previously been reinforced (Garon et al., 2008). The inhibition assessed in Isquith et al.'s (2004) model is response inhibition.

Both forms of inhibition potentially play a role in language acquisition. Resistance to interference is thought to play an indirect role in developing language competence by allowing efficient deployment of total mental capacity (Im-Bolter, Johnson, & Pascual-Leone, 2006). It could also play a role in discourse, being used to disregard alternative interpretations of ambiguous utterances (Im-Bolter et al., 2006; Viterbori, Gandolfi, & Usai, 2012). Response inhibition could play an important role in corrective learning, as previously used mental schemes can only be replaced with updated ones if the old scheme is not being constantly activated. For example, in order to progress past use of overgeneralisations (such as "Dad" to refer to all people), a child must begin to inhibit this scheme and instead activate specific names for individuals.

Studies investigating the link between inhibition and language skills have found equivocal results in school aged children. Two studies have shown children with SLI to have an increased difficulty with interference relative to TD and or younger language-matched peers, but not with response inhibition (Marton et al., 2012; Marton and Schwartz (2003)). However other studies have shown poorer response inhibition in the same population (Bishop & Norbury, 2005b; Im-Bolter et al., 2006; Oram, 2003; Spaulding, 2010). Yet Bishop and Norbury's (2005a) conclusion was that it was most likely poor language or comorbid attention difficulties which caused the poor inhibition performance rather than vice versa in their second study, as children with autism with similar language and attention problems showed similar levels of performance. Noterdaeme and colleagues (2001) and Henry et al. (2012) found mixed results, with individuals with SLI (primary and secondary school age) performing significantly poorly on one response inhibition task, but not on the other. There are many possible reasons for these conflicting findings, including the heterogeneity of the

SLI population, small sample sizes, differing developmental trajectories for interference compared with response inhibition and methodological differences between studies.

The picture seems clearer at a younger age. Conboy, Sommerville and Kuhl's (2008) study indicated that response inhibition skills may play a role in a child's progression to neural commitment for the native sound contrasts. They found that 11 month old infants' scores on cognitive control tasks measuring response inhibition correlated with the children's ability to ignore non-native speech sound contrasts. This would presumably predict better language outcomes for those who progressed to neural commitments faster, although this has not been tested longitudinally. Viterbori et al. (2012) found moderate associations between response inhibition and morphosyntax skills in a study of TD Italian speaking preschoolers aged 24-36 months. In addition, TD preschoolers' scores on inhibition tasks were found to correlate with receptive vocabulary measures to a small degree at ages three-to-four years (Epsy et al., 1999; Willoughby et al., 2010). These studies did not note the presence or absence of late talkers in their samples. The relationship between inhibition (as measured by the BRIEF-P) and language has been shown to be equivocal. A sample of 21 preschool aged children with LI showed a significantly higher level of problems with Inhibition on the BRIEF-P, than their age-matched peers (Gioia et al., 2003). However, Wittke, Spaulding, and Schechtman (2013) found no group differences in Inhibition in their sample of 19 SLI preschool aged children and 19 age- and gender-matched controls.

To summarise, the literature generally supports the concept of a small to moderate association of response inhibition and language in the preschool years, but the literature is less clear regarding such associations in school aged children. Positive associations with interference inhibition and language are more commonly found at school age; this aspect of inhibition cannot be measured in toddlers as it has not yet developed (Marton et al., 2012).

Response inhibition may be a useful predictor of language outcomes in longitudinal studies from toddlerhood.

2.6.3.3 Shifting

Shifting refers to the ability of the individual to change between activation of conflicting mental schemes for successful task completion. Tasks measuring shifting typically require alternating between two newly learned rules. For example, in card sorting tasks, the pictures on the cards have different colours and shapes. At first the individual is required to sort by colour, then to disregard colour and sort by shape. Individuals must retain the rule in working memory and inhibit the first mental scheme while activating the new one. This demonstrates how shifting builds on the EFs of working memory and inhibition (Garon et al., 2008).

Shifting could conceivably play a small role in the language development. While it is unlikely that shifting is a major mechanism of language acquisition, a better ability to inhibit old patterns of verbal processing and activate a new response set, should aid the child in moving to more adult language forms. This would especially be the case where the child was aware of their errors.

Difficulties with shifting have been found in some studies of primary school aged children with SLI (Marton, 2008), but not in others (Henry et al., 2012; Weyandt & Willis, 1994). Im-Bolter et al. (2006) studied the EFs of children with SLI compared with TD controls. While shifting and language competence showed a moderate correlation, this study did not support a direct role for shifting in language competence. Path analysis showed that the model where shifting was linked to inhibition and working memory capacity, which in turn correlated with language, was the best fit. They concluded that shifting was not important for language development, but that language can be used to support such shifts. Noterdaeme et al. (2001) came to a similar conclusion in their study of children with SLI. In contrast Dibbets, Bakker, and Jolles (2006) reported that despite no group differences in shift task performance compared with a control group, the children with SLI (*n* = 6) showed increased recruitment in cortical areas relating to executive control during shift tasks. They interpreted this to mean that the children with SLI found the task more difficult than controls, even although their performance was the same. This finding awaits replication with a larger sample. At a preschool level, Willoughby et al. (2010) found their shifting task was moderately correlated with receptive vocabulary in their large community study of three year olds. Both studies (previously mentioned) which used the BRIEF-P to measure EF in pre-schoolers with LI found increased reported problem behaviours with Shift relative to age-matched peers (Gioia et al., 2003; Wittke et al., 2013).

Overall, shifting seems to be associated with LI at the preschool level, even if the picture is not so clear at school age. It would seem more likely that language development facilitated better shifting than vice versa. However, the evidence is not clear. Shifting may be a useful predictive factor for language in longitudinal studies of toddlers.

2.6.3.4 Metacognitive executive functions

The final two components of Isquith's model (Emotional Control and Plan / Organising) are metacognitive functions, rather than basic EFs, and as such are more likely to require language for their development than vice versa. These EFs build on the other three EFs in the model; for example, inhibition could be used to restrain an emotional outburst, shifting to choose an alternative to habitual organising behaviours, or working memory to support planning strategies. As the BRIEF-P includes these final two EFs, data will be readily available on them. Therefore they have been included in the current research as a point of comparison.

2.6.3.5 Plan / Organise

"Plan / organise" is the ability to manage current and future task demands in context. Planning is the ability to move towards a goal by strategically executing steps to reach that goal, often in a necessarily sequential order. Organising refers to the ability to bring order to actions, materials or information to achieve a goal (Isquith et al., 2004). Planning / organising could be involved in language acquisition by allowing better formulation of discourse level interaction and narratives. While discourse and narratives begin at a very simple level, these exchanges still require planning in the first instances.

The literature does not provide a definitive answer to whether poor planning / organising abilities are associated with language acquisition. Primary school aged children with SLI have been shown to have poorer planning skills than their TD peers in EF tasks (Henry et al., 2012; Marton, 2008). Henry et al. (2012) found that these group differences remained significant even when age, non-verbal and verbal IQ were controlled. This indicates that children with SLI may have difficulties with planning that cannot be accounted for by their lower language skills. Yet Kamhi (1995) found no evidence of planning difficulties in his five-to-seven-year-old sample of boys with SLI. He concluded that poor planning did not impact on language skills at this age, but suggested that before language abilities become automatic and modular, poor planning skills may impact on language development. Epsy et al. (2001) found moderate correlations between planning task scores and receptive vocabulary in 30-60 month old TD children. The two BRIEF-P studies show conflicting results. Gioia et al. (2003) found that their TD and LI groups scored similarly in Plan / Organise, whereas Wittke et al. (2013) with a similar sample and the same measure reported significantly poorer Plan / Organise scores in their LI group.

In summary, equivocal results in the literature likely point to heterogeneity in the LI population, as there is a trend for smaller studies (n = 15-21) to fail to find a significant relationship compared with studies using a larger sample size (n > 40). In terms of a casual direction, it does seem more likely that better language impacts on better planning / organising than vice versa. However, early Plan / Organise scores may be predictive of language outcomes in the age range two to four years.

2.6.3.6 Emotional Control

Emotional control is considered to be the regulation of one's emotions to enable successful social interaction in a variety of situations. Children start with being emotionally regulated externally by their parents (receiving comfort or encouragement) and gradually shift to internal regulation. Internal language is increasingly used to mediate this process (e.g. "it's okay, it will be over soon") (Isquith et al., 2004).

Late talkers and preschoolers with LI have been shown to have higher instances of behavioural disturbances than TD children, and to have more difficulties with emotional control (Gioia et al., 2003; Irwin, Carter, & Briggs-Gowan, 2002; McCabe, 2005; Wittke et al., 2013). In school aged children, Fujiki, Brinton, and Clarke (2002) found that boys with SLI had more difficulties with emotional regulation than TD children, and that this worsened the older the child. However, in this study emotional regulation was not correlated with language scores. Redmond and Rice (1998) found evidence to suggest that once language difficulties had resolved, the social-emotional difficulties of children with SLI also resolved, suggesting social difficulties are a result of poor language rather than a cause in most cases. However, some have suggested the causality between emotional control and language is bidirectional, and varies from individual to individual (Durkin & Conti-Ramsden, 2010). It seems more likely that better language development supports better emotional control, but it is also possible that basic EFs (inhibition, working memory, and shifting) may dually impact on emotional control and language, thus causing an association.

Difficulties with emotional control could arise from a range of factors, such as a weakness in the CE, mismatch between parenting style and a child's temperament and difficulty with communication associated with early language delay. This relationship is likely to be a complex one and may vary from individual to individual. While in this study, emotional control will be investigated primarily as a measure of the CE, the complexity of this construct will be considered when interpreting the results.

2.7 Summary of literature review

While most late talking toddlers catch up with their peers in time, many parents seek professional advice early on. Distinguishing which toddlers will catch up and which ones will not is problematic. Understanding risk factors for both early expressive delays and later LIs would be informative on both clinical and theoretical levels. Despite the sizable body of research on late talkers to date, predictive models remain moderately successful at best. I have argued that successful prediction of language outcomes requires inclusion of psycholinguistic processing skills in predictive models. Domain general theories of SLI predict that deficits in lower level cognitive processing will likely result in difficulty learning language. As deficits in working memory relate to three of the four causal theories, the role of working memory in early language acquisition was chosen as the focus of this study. The role that each of the following aspects of working memory might play in language acquisition was discussed from different theoretical viewpoints: phonological loop; processing speed and the CE (considered from three angles: VWM, VSWM and EF). A review of the literature showed that impairments in these aspects of working memory have been implicated in language variability in children and young adults in a variety of clinical and typical populations. Yet very few studies have investigated these variables with regards to two-year-old language variability, and whether these aspects of working memory could aid longitudinal predictive models for language in age range two to four years. If working memory plays a role in language acquisition, it is most likely to be detected in the age range two to four years, when a substantial proportion of language learning occurs. There are two overarching hypotheses for research. Firstly, deficits in aspects of the working memory system are implicated in late talking. Secondly adding working memory measures to existing predictive models will improve predictions of language outcomes across the age range two to four years. These hypotheses are made in accordance with the theory that LIs are caused by capacity limits in processing and the view that late talkers represent the lower end of the spectrum of ability in language.

2.8 Overview of the current research

Three studies are presented in this thesis. Study 1 examined the concurrent relationships between measures of the working memory system, visual cognition and language in TD and late talking in children aged 24-30 months. Group comparisons between the late talkers and TD groups were made to see whether aspects of working memory were implicated in early language delays. Study 2 repeated this methodology after an 18 month period to examine the stability of relationships between variables over time. The RLTs at 42-48 months were compared with the TD group to see whether aspects of working memory could account for their pattern of early delay followed by rapid resolution of language skills. Study 3 reported the clinical outcomes of the cohort. It examined whether aspects of working memory measured at 24-30 months improved known predictive models for language outcomes at 42-48 months of age in this cohort. Finally the issue of individual prediction, which is the overall goal of this research, was considered. The research questions and hypotheses for each study are as follows.

Study 1

Research questions

1. What are the bivariate and multivariate associations between the expressive vocabulary skills of 24-30 month olds and the following aspects of the working memory system: PSTM, processing speed, VWM, VSWM and EF (Inhibition, Shift, Emotional Control, Working Memory, Plan / Organise)?
2. Do these aspects of working memory improve on previously established multivariate linear regression models for predicting expressive vocabulary at 24-30 months?
3. What differences in group means can be seen between the TD and late talking groups in the above aspects of working memory?
4. Do any significant differences remain when language differences between the two groups are controlled?

Research hypotheses

1. Significant associations will be seen between all aspects of the working memory system and expressive vocabulary on a bivariate level. PSTM is expected to show the strongest association with expressive vocabulary in the multivariate model.
2. Working memory measures will improve previously established multivariate linear regression models predicting expressive vocabulary at ages 24-30 months.
3. Performance on measures of working memory will be lower in late talkers than in TD children, particularly in PSTM.
4. These effects are expected to be seen even when controlling for group differences in language skills.

Study 2

Research questions

5. What are the bivariate and multivariate associations between the expressive language skills of 42-48 month olds and the following aspects of the working memory system: PSTM, processing speed, VWM, VSWM and EF (Inhibition, Shift, Emotional Control, Working Memory, Plan / Organise?
6. Do these aspects of working memory improve on previously established multivariate linear regression models for predicting expressive language at 42-48 months?
7. What differences in group means can be seen between the RLT and TD groups in working memory and language at ages 42-48 months?
8. Do any significant differences remain when language differences between the two groups are controlled?

Research hypotheses

5. Significant associations will be seen between all aspects of the working memory system and expressive language on a bivariate level. PSTM is expected to show the strongest association with expressive vocabulary in the multivariate model.
6. Working memory measures will improve previously established multivariate linear regression models predicting expressive vocabulary at ages 42-48 months.
7. Performance on measures of working memory will be lower in RLTs than in TD children, particularly in PSTM.
8. These effects are expected to be seen even when controlling for group differences in language skills.

Study 3

Research questions

9. What clinical outcomes are evident at ages 42-48 months?
10. What are the bivariate and multivariate associations between the total language score on the PLS-4 at 42-48 month olds and the following aspects of the working memory system: PSTM, processing speed, VWM, VSWM and EF (Inhibition, Shift, Emotional Control, Working Memory, Plan / Organise) measured at 24-30 months?
11. Do these aspects of working memory improve on previously established multivariate linear regression models for predicting later language at 42-48 months of age?
12. Do working memory measures at 24-30 months improve prediction of total language outcomes at 42-48 months for individual late talkers?

Research hypotheses

9. 50%-75% of late talkers are expected to resolve and score within the normal range on total language scores on the PLS-4 by ages 42-48 months. Due to the liberal criteria for late talker status, no former TD children are expected to have a LI.
10. Significant associations will be seen between all aspects of early working memory skills and outcome total language scores on a bivariate level. Working memory skills will predict unique variance in the multivariate model, the strongest predictor being PSTM.
11. Aspects of working memory measured at 24-30 months are expected to improve established predictive models of total language outcomes at 42-48 months.
12. The addition of early working memory variables into predictive models is expected to improve prediction for individual late talkers.

Chapter Three: Study 1 Methodology

3.1 Recruitment

3.2 Participants

3.3 Measures

3.3.1 Questionnaires

3.3.2 In-clinic behavioural measures

3.4 Procedures

3.1 Recruitment

Ethical permission to conduct the study was gained from the Human Ethics Committee at the University of Canterbury (see Appendix A).

Recruitment was conducted in a two-step process. Part 1 of the study aimed to gather preliminary information about the children's vocabulary level and their demographics. Data from children who only completed Part 1 of the study have not been included in the data analysis. This part served as a gateway to recruitment for Part 2 of the study. Parents of all children aged 24-30 months were invited to participate in Part 1. Part 1 participants were recruited through personal networks, a university research database, doctors' offices, local special education providers (the Ministry of Education and a centre for children with developmental difficulties), Plunkett nurses (the local preschool public health service) and early childhood education centres. On receiving information about the study, interested parents contacted the Child Language Centre for a parent pack (containing the MacArthur-Bates CDI, Parent Questionnaire I, and the study information sheet and consent form). See Appendix B. Plunkett nurses also handed out parent packs directly to parents with children in the target age range. Parents then filled in the questionnaires and sent them to the Child Language Centre by return post. This completed Part 1 of the study.

Parents who were interested in participating in Part 2 indicated this in the Parent Questionnaire. Potential participants were screened to check they met criteria for Part 2. There were three exclusion criteria: Children could not have a diagnosis known to affect speech, hearing or language development. Secondly, children were required to have English as their main language. If parents indicated their child was exposed to languages other than English, they were contacted to determine the level of exposure. If the exposure to other languages was reported by the parent to be less than 20% of the time, they were invited to participate in Part 2 as outlined below. It was decided not to exclude children with some exposure to other languages in order to keep the level of participation from ethnic minorities as high as possible, so that the sample was more representative of the New Zealand community. However, fully bilingual children were excluded. Finally children had to live within 100km of Christchurch so they were able to attend assessment appointments.

In order to gain the sample size needed for the required statistical analyses, particular effort was made to recruit late talkers into the study. Once recruitment was underway, parents already in Part Two were asked if they knew any children in this age range who were late to talk and if they would be willing to forward information about the study to their parents. Towards the end of the recruitment period, preschool managers were also informed via email that I particularly needed more late talkers, and they also may have encouraged parents of such children to participate in this research. In addition, in an attempt to better balance the sample, towards the end of the recruitment period some TD children of high SES status were turned away from Part 2 in favour of TD children with lower SES status.

Parents of children who met the requirements of the study were contacted regarding Part 2 of the study with a second information sheet (see Appendix C). Participation in Part 2 involved a commitment to complete a protocol of assessments with their child at two time points which were 18 months apart. The initial assessment required two visits to the Child Language Centre while their child was aged 24-30 months. Outcome assessments would require a further two visits in 18 months' time when their child was aged 42-48 months. This completed the families' involvement in the study.

A $20 voucher to a local shopping mall and a children's book were offered as incentives for attending the initial assessments at ages 24-30 months. Another $20 voucher and a petrol voucher for $10-20 (depending on whether the family lived within city limits or had travelled further) was offered for completing the outcome assessments at ages 42-48 months. Signed parental consent was gained for participation in Part 2 when the child and parent made their first visit to the Child Language Centre (see Appendix C for a copy of the consent form for Part 2).

A total of 82 children were recruited to Part 2 of the study. Of these, two children were assumed to have withdrawn from the study before completing the initial assessments (due to failure to attend appointments and a lack of response to attempts to contact them), and one was excluded from the analysis as she had completed less than half of the assessment protocol after four appointments. Data from the remaining children were included in the data analysis.

3.2 Participants

The participants were 79 children aged 24-30 months. There were more boys (68%, n = 54) than girls in the sample, due to the disproportionate numbers of late talkers in the cohort. Most of the children were first- (58%, n = 46) or second-born (32%, n = 25). One child was a twin (her twin was not a participant in the study). English was the main language of all the households, however, 17% (n = 13) reported using a second language at home less than 20% of the time (including NZ Sign Language). There was a high incidence of reported positive family history for speech, language or learning difficulties (43%, n = 34). This was coded as a “yes” response if the parents listed a first- or second-degree relative. Types of difficulties reported ranged from difficulties with articulating /s/ to global developmental delays. There was also a high level of parent concern about language (22%, n = 17), which was likely due to the inflated number of late talkers in this study.

Table 3.1 summarises the demographic information from the Parent Questionnaire I which was completed when the children were 24-30 months of age.

Table 3.1

Summary of parent and child demographic characteristics when the children were aged 24-30 months

	Total Sample (N = 79)		
Variable	% of sample (N)	M (SD)	Range
Parent Characteristics			
Highest qualification achieved:			
<11 years (no qualifications)	3 (2)		
11-12 years (NCEA* L1-2) (junior high school)	9 (7)		
13-14 years (NCEA* L3-4 or trade certificates) (senior high school)	14 (11)		
Diplomas or advanced trade certificates	12 (9)		
Bachelor's degree	40 (31)		
Honours, master's, postgraduate diploma, PhD	23 (18)		
Not specified (overseas secondary school qualification)	0 (0)		
Not specified (other high school qualification gained in New Zealand)	1 (1)		
Parent concern for child's:			
Hearing	6 (5)		
Language	22 (21.5)		
Communication	11 (9)		
Child Characteristics			
Age (months)		26.62 (2.04)	24-30
Gender (boy)	68 (54)		
Birth order:			
1	58 (46)		
2	32 (25)		
3	6 (5)		
4	4 (3)		
5			
Number of children in the family:			
1	32 (25)		
2	54 (43)		
3	10 (8)		
4	1 (1)		
5	3 (2)		

Variable	Total Sample (N = 79) % of sample (N)	M (SD)	Range
Birth weight (g)		3528.35 (539.05)	1340-4850
No major health problems	89 (70)		
Prematurity (weeks)		.58 (1.53)	0-10
1	6 (5)		
2	4 (3)		
3	5 (4)		
4	3 (2)		
5	1 (1)		
10	1 (1)		
Other language spoken at home	17 (13)		
Hours per week in day care		12.95 (12.21)	0-50
Family history of speech, language or learning difficulties	30 (24)		
Hearing (OAE screen at 2000, 2500, 3200 and 4000 hz at 50 dB):			
"Pass" one ear and "refer" for the other	20 (16)		
"Pass" both ears	65 (51)		
"Refer" one ear and no data for the other	5 (4)		
Noncompliant (both ears)	10 (8)		

Note. NCEA = National Certificate of Educational Achievement ; OAE = Otoacoustic Emissions

The ethnic balance of the sample was reasonably representative of the Canterbury region as can be seen in Table 3.2, with the exception of a slightly over represented New Zealand (NZ) European group.

Table 3.2

Comparison of the sample and population ethnic mix

Ethnicity	% of sample (N)	% of total population in Canterbury region[3]	% of total population in New Zealand
NZ European	85 (67)	75	65
Maori	8 (6)	7	14
Asian	4 (3)	6	9

[3] NZ Census information 2006 downloaded from: http://www.stats.govt.nz/Census/2006CensusHomePage/QuickStats.aspx

Ethnicity	% of sample (N)	% of total population in Canterbury region[3]	% of total population in New Zealand
Pacific Islands	3 (2)	2	7
Middle Eastern / Latin American / African	0 (0)	1	
Not specified ("New Zealander")	1 (1)	13	11

Note. NZ European = New Zealand European.

As can be seen in Table 3.3, the education level of the parents in the sample was higher than that of the population of adults aged 20-44 years in Canterbury (StatisticsNZ, 2006), with a higher percentage holding university degrees and a lower percentage with no or few qualifications. This is not unexpected in a self-selected sample, but it does mean that the cohort may not be representative of the general population. In most cases (76/79) the mother of the child completed the form, and therefore it was her highest level of education which was recorded. However for three children, it was the fathers' highest qualification level instead.

Table 3.3

Highest parent qualifications in the sample compared with population frequencies

Highest parent qualification	% of sample (N)	% of Canterbury population (ages 20-44 years) [4]
<11 years (No qualifications)	3 (2)	15
11-12 years (NCEA L1-2) (Junior High School)	9 (7)	25
13-14 years (NCEA L3-4) (Senior High School) / Trade Certificates	14 (11)	22
Diplomas or Advanced Trade Certificates	12 (9)	10
Bachelor's Degree	40 (31)	14
Honours, Postgraduate Diploma, Masters, PhD	23 (18)	5
Not specified (Overseas secondary school qualification)	0 (0)	5
Not specified (Other NZ high school qualification)	1 (1)	6

Note. NCEA= National Certificate of Educational Achievement; L1-4 = Levels 1-4.

[4] NZ Census information 2006 downloaded from: http://www.stats.govt.nz/Census/2006CensusHomePage/QuickStats.aspx

3.3 Measures

A battery of 11 assessments was used in the study, as outlined below. The current study was a part of a larger study ("Learning to Talk"), the protocol of which included two additional measures, a parent concern questionnaire (Parent Evaluation of Developmental Status (PEDS); (Centre for Child Community Health, 2006)) and a language sample. The results of these two assessments are not been reported here, as they are components of the wider Learning to Talk study only. Each of the 11 measures used in the current study is described below. The scores forms for the non-standardised measures are included in Appendix D.

3.3.1 Questionnaires

1. **MacArthur Bates - Communicative Development Inventory (CDI), New Zealand English Adaptation (Reese & Read, 2000)**

The CDI: Words and Sentences (Fenson et al., 2007) is a well-established parent report measure of expressive vocabulary and grammatical development for children aged 16-30 months. Parents completed this checklist of 680 words and questions about language use and returned it to the research team. The CDI provides a range of scores on various aspects of expressive language development, however the only measure used in the statistical analyses in the current research was *total words produced.* The CDI has high internal consistency, test-retest reliability, good concurrent and predictive validity (Fenson et al., 2007; Reese & Read, 2000) and has been normed from a large sample in the United States of America.

2. **Parent Questionnaire I**

This questionnaire was developed by the research team running the overall "Learning To Talk" study (Thomas Klee, Stephanie Stokes, Catherine Moran and Jayne Newbury). Wording and format was based on validated questionnaires for the NZ population, such as the

NZ 2006 census survey questions. It gathered information about demographic variables and parent concern about their child's language development.

3. **Behaviour Rating Inventory of Executive Function - Preschool Version (BRIEF-P)**

The BRIEF–P (Gioia et al., 2003) is a parent report measure of EF in everyday settings. It contains a checklist of 63 items. Parents indicate if a listed behaviour has been a problem 'often', 'sometimes' or 'never' in the last six months. A *Global Executive Composite (GEC)* is gained from adding up the responses to all items. In this study, five indices were also calculated: *Inhibition, Shift, Working Memory, Emotional Control and Plan / Organise.* The scores of these indices were tallied from the items in the test which corresponded to these constructs. The BRIEF-P has high internal consistency, test-retest reliability and concurrent validity (Gioia et al., 2003). It has been shown to distinguish between TD and clinical samples of children well. Inter-rater reliability tends to be modest, particularly if the child is observed across different settings (e.g. home and school). This is because the perception of the rater is an important determiner of the scores given and the differences in children's behaviour across social contexts. For this reason the BRIEF-P is not intended to stand alone as a measure of EF, but is intended to be used as part of a wider battery of EF assessments as an additional perspective on the child's development (Isquith et al., 2005).

3.3.2 In-clinic behavioural measures:

1. **The Mullen Early Scales of Learning: Visual Reception Organisation (VRO)**

The Mullen Early Scales of Learning (Mullen, 1995) is a comprehensive assessment of children's early cognitive, language, motor and sensory development. Only the Visual Reception Organisation (VRO) subscale was administered in this study. This subtest was chosen as an indicator of visual cognitive ability because of its suitability to the age of the children in this study and its time efficiency. Test item format is varied but includes matching, categorising and recalling visual or spatial information. This test provides both *raw scores* and *standard scores*. The internal consistency of this subscale is high in the age range two-to-four years and test-retest and inter-rater reliability is also high. Concurrent validity was reported as good in the test manual (Mullen, 1995).

2. The Preschool Language Scale: Fourth Edition (PLS-4) (Australian Language Adapted)

The PLS-4 (Zimmerman, Steiner, & Pond, 2002) is a multiphasic assessment of children's language abilities. It was chosen because it covers emergent communication skills through to language skills needed for academic tasks (age six years, eleven months), so was well suited to capture the wide variation in language development in this cohort. The PLS-4 was the only available multiphasic language measure which could have been used for both age ranges tested in this research. The PLS-4 provides two subscale scores; the Auditory Comprehension (PLS-4 AC) and Expressive Communication (PLS-4 EC) subscales. Each subscale includes items which cover lexical, morphological and syntactic knowledge. Item testing format varies, but typical items involve pointing to the correct picture in response to a question, answering a question verbally, following an instruction either in a natural situation or with a toy array. This test provides *raw scores* and *standard scores* for the AC and EC subtests and also for total language ability. Test-retest reliability, internal consistency and inter-rater reliability were reported as high. Concurrent validity was reported as good in the test manual. The PLS-4 has been shown to be effective in detecting children with language impairments (Zimmerman et al., 2002).

3. The Receptive One Word Picture Vocabulary Test (Fourth Edition) (ROWPVT-4)

The ROWPVT-4 (Martin & Brownell, 2011) was included as a measure of receptive vocabulary. Children are shown a page of four pictures and asked to point to a named vocabulary item. Because this test was developed for American children, two changes were made in accordance with NZ English vocabulary ("cookies" to "biscuits" and "mailman" to "postman / postie"). One target word was also changed (from "baseball" to "rugby"), as baseball is not a common sport in NZ. The manual reports that internal consistency and test-retest reliability is high. This test provides both *raw scores* and *standard scores*. Concurrent validity was reported to be moderate. The ROWPVT-4 distinguishes between typical and clinical populations at a group level (Martin & Brownell, 2011).

4. The Toddler Phonology Test (TPT)

The TPT (McIntosh & Dodd, 2011) was chosen as the measure of phonology as it was designed for children aged 2;0-2;11 years. Children were asked to name 31 pictures targeting 37 words. The words cover 105 consonants in syllable initial and syllable final positions, and 56 vowels and diphthongs. A range of syllable structures are covered. Broad transcription of the child's speech production was made and analysed for phonological and articulation errors. The TPT has developmental norms from Australia and the United Kingdom. Test-retest and inter-rater reliability were reported to be high in the test manual (McIntosh & Dodd, 2011). Only one metric from this assessment was used in the statistical analyses: *Percent Consonants Correct (PCC).* The PCC was calculated according to the instructions in the manual, as the percentage of phonemes correct in words attempted by the child. Some late talkers did not have enough expressive vocabulary to complete this test, even by imitation of words. There was no guidance for how to score these children in the test manual, so the following protocol was established for this study. Children were presented with the task and if they were unable to name or imitate words in the test, their phonology score was entered as zero. This was only done if the CDI and discussion with parents confirmed that this behaviour was representative of their expressive language level. If there was any doubt, the score was entered as missing data. Other children completed part of the test but were unable to attempt every word due to limited expressive language. Those who attempted nine words or more had their PCC on attempted words calculated and entered alongside those who completed the test. Nine words seemed to be a natural cut-off as there were nine children who attempted more than nine words but did not complete the test (average number of words attempted in this group was 24/37); no children who only attempted two to eight words; and then only two children who just said one word. All the children who attempted fewer than nine words on the TPT were entered as a zero score.

PCC does not distinguish between types of error (e.g. developmentally appropriate errors cf. delayed or disordered sounds), but instead serves as a summary of the child's phonological development. However a qualitative analysis of the number and type of speech errors relative to the child's age was incorporated in any clinical judgements of concern for phonology.

5. The A not B task

The A not B task was originally developed by Piaget (1954). It was included in this study as a measure of the CE in the visual-spatial domain. It is considered to measure mainly VSWM, although inhibition and shifting could be seen to be involved on reversal trials. Children are shown two identical cups 2 cm apart. Their attention is alerted to a motivating reward (usually food), which is then hidden under one of the cups. The cups are lowered over the reward simultaneously. The child's view of the cups is then blocked by a screen for a varying amounts of time (5, 10 or 15 s). The screen is then removed and the child asked "where is it?" If the child finds the reward, they can eat it, if not, it is hidden again in the next trial. Self-corrected reaches to the incorrect cup were scored as correct. When the reward type was no longer motivating, it was changed to a preferred option (e.g. another type of food, sticker or toy). This is critical to ensure the child's motivation for the task (Epsy et al., 2001). The side that the reward was hidden on was switched after two consecutive correct searches.

The trial on which the side changed is known as a “reversal trial”. This trial is the one where the CE is most taxed, as the child must inhibit a previously rewarded motoric / mental schema to find the reward on the new side.

The following administration procedures were based on procedures described by Griffith et al. (1999) and Diamond (1990), and were developed through a pilot study. Testing began at 10 s delay. If the child scored an error in the first three trials, the delay was reduced to 5 s. If there were no errors in the first three trials at this level, testing resumed at a 10 s delay until 10 trials had been administered at this level. If there was an error on the first three trials at 5 s delay, a further seven trials were administered at this level. If there were no errors in the first three trials at the 10 s delay, the delay was increased to 15 s, and 10 trials were administered at this level. The aim of this was to find the shortest delay at which an error could be elicited. This protocol was designed to reduce the number of trials the child was asked to complete. As this assessment relies on motivation and attention, this was a critical consideration. The *total score* was calculated by totalling the number of trials correct at the child’s delay level, plus 10 points for a 10 s delay and plus 20 points for a 15 s delay level. This meant the possible range of total scores was 0-30. While the *delay level* at which errors were elicited was also available, only the *total score* was used in the statistical analyses for the current research.

Developmental progression in the A not B task has been demonstrated from infancy through to age five years, with a reduction of a ceiling effect at ages four to five years expected with delays of more than 10 s (Epsy et al., 2001). Performance on the A not B task has been linked to the dorsolateral prefrontal cortex in monkeys, which confirms the involvement of the CE (Diamond, 1990). A very similar task, Delayed Response has achieved poor test-retest reliability in the age range 2;6-6;0 years, r (33) = .44 (Epsy, Bull, Kaiser, Martin, & Banet, 2008). The other options for measuring VSWM in this age range had lower test-retest reliability (e.g. the Six Boxes task, r (33) = .33 (Epsy et al., 2008)). Overall this task seems to be valid and sensitive to developmental changes in this age range, but its reliability is reported to be less than ideal.

6. The Key Word Working Memory I task (KWM I task)

The Key Word Working Memory task (KWM I) was included as a measure of VWM in the current research. This task was developed by the "Learning to Talk" research team through a piloting process. Children were asked to follow a series of instructions which differed by the number of key words they had to recall. Use of VWM (storage and processing) was indicated by the child acting out the instruction using a set of toys. The processing component was demonstrated by the child's association of the spoken words to the toys / actions required. Storage was assumed if the child followed the instruction (as for the noisy book task described in Hughes (1998)).

Three items were administered at Key Word Levels one to four (1-4 KWLs). An example of a 1KWL instruction is "where's the *pig*?" An example of a 4KWL instruction is "*pig* wants the *bed* and *dog* wants the *drink*". The child was credited with processing and storing a key word if they acted on a named toy(s) in response to the instruction. Partial credit was allowed (for example if the instruction was "*cat* wants the *bed*" and the child put the cat in the bath, the score was one out of a possible two points for that item). Testing continued until the child had completed all 12 items, refused to cooperate any further, or missed at least two key words on each instruction at a KWL. The set of toys was: pig, dog, cat, bed, bath and drink. If the child did not know the vocabulary for "cat", "dog" or "pig", a male or female doll were substituted (referred to as "dad" and "mum"). If they did not know the vocabulary for "drink", "bath" or "bed", a toy plate ("food") or book ("book") were substituted. The verb "wants" was expected to be understood in this context by all the children (74% of children aged 8-18 months were reported to understand this verb by their parents in the "CDI: Words and Gestures" normative sample (Dale & Fenson, 1996)). Due to the developmental level of the children, if they missed an instruction due to inattention or non-compliance, an alternative was administered. Also, unconventional responses were accepted if it was clear the child did recall and comprehend the instruction for example, if the child made the pig drink from the bath tap in response to "pig wants a drink", this was scored as correct. The child was then shown how to make the pig drink from the cup and discouraged from letting them drink from the bath taps. Such responses were uncommon so this was not a concern.

Two scores were calculated from this task: The *total score* was determined by adding up the total number of key words recalled across items attempted. The *KWL* was the highest item score that the child achieved at least twice throughout the test. Scores were determined online, and then checked on the video for accuracy. Only the *total score* was used in the statistical analyses of the current research. See Appendix D for the score form.

The construct validity of the KWM I task was carefully considered in the development of its administration and scoring procedures. The KWM I task does require linguistic understanding. By definition, it must, to measure *verbal* working memory. In other words, children must understand some language in order to have VWM capacity. However the increases in difficulty across KWLs increase demands on the memory system rather the linguistic system, as children were not scored incorrectly for recalling key words out of order. For example if the instruction was "the pig wants the bed and the cat wants the bath", a child who gave the cat the bed and the pig the bath was still scored correctly. It was therefore primarily memory rather than knowledge of compound sentence structures and word order which were being tested. Vocabulary was also carefully checked to ensure the child was familiar with all the words in the instructions before beginning the assessment.

Inter-rater reliability using the recordings was examined. Twenty of the 79 participants were rescored by a research assistant from video recordings. The first 20 recordings with a good quality audio / visual recording of the task were selected. Inter-rater agreement was 97%. The total raw score was the same or differed by one point 19/20 times. It was exactly the same 45% of the time, r (18) = .95. Discrepancies could be grouped into three categories: scoring error, discrepancy in interpreting scoring rules, disagreements in how to score unusual responses from children (e.g. is giving the pig a drink in the bath acceptable for "pig wants a drink"? Or should a response which looks like play, rather than a response to the instruction, be scored?). Eighty-five percent of the KWLs were the same across both scorers and all were within one level of each other. A slightly greater inter-rater reliability could have been achieved by making the scoring rules less flexible, to reduce the amount of subjective interpretation required by the scorer.

7. **The Looking While Listening task I (LWL I task)**

The Looking While Listening I (LWL I) task was developed and administered broadly following the methodology outlined in Fernald, Zangl, Portillo, and Marchman (2008). The LWL I task serves as a measure of processing speed, as it results in a mean speed of spoken word recognition for each child. The child's task was to look at two pictures on screen (a target and a distractor), listen to the instruction "look at the <target>" and respond by looking at the target picture. On trials where the child was looking at the distractor picture initially, the time taken from the start of the target word to when the child initiated an eye gaze shift to the target picture was calculated and termed a "latency to shift." An average of latencies to shift was calculated for each child and was referred to as a "*mean latency*" in this research. Trials where the child was already looking at the target picture were discarded.

Task presentation format

Each child was seated in front of a computer monitor, with a camera mounted above, and focused to gain a head and shoulders shot. Children were usually seated on their parent's lap, although a few preferred to sit alone. Parents were given blacked-out sunglasses to wear during the task so they could not see the stimuli, or if the child objected to this, the parent looked down at their child's back throughout the task. The children then watched a PowerPoint presentation which was five minutes in length. Their eye movements were recorded for later video analysis.

The PowerPoint presentation consisted of 40 trial slides. Each trial slide had two photograph images on it. One represented the target word and the other the distractor. All pictures were matched for size, colour, salience and interest by visual comparison. Animate objects were always paired with animates and vice versa. Word pairs always began with a different phoneme. As much as possible, each image served equally as target and distractor, and appeared equally on the left and right. In the original version of the task (first 10 participants), the target image appeared on the left and right an equal number of times, and did not appear on any side consecutively for more than three trials in a row. After seeing the two images on a trial slide for two seconds, the pre-recorded sentence "where's the <target>?" was played. There was then another two seconds before the images disappeared and the trial ended. There was a blank slide between each trial for one second, meaning each trial was approximately six seconds long. Four trial slides were played, followed by a reinforcer slide of a favourite cartoon character to help maintain the children's interest. Short tag questions (such as "can you see it?"; "show me which one"; "point to it") were played after the stimulus phrase on each trial in the original version of the task. These were included as recommended by Fernald et al. (2008) to help keep the children's interest in the task by adding variety. All audio recordings were made by a female speaker of NZ English on a Zoom H4 Handy Recorder in stereo at a sampling rate of 96kHz.

The word list was chosen from the CLEX data base (Dale & Fenson, 1996). Half of the items in the word list had been acquired by age 18-23 months (shoe, banana, phone, chair, juice, cat, dog, baby, horse, car, tree), and half by age 24-30 months (glass, toast, clock, star, fish, sheep, elephant, zip, pig and cow). “Acquired” was defined as words which 75% of the children in this sample had said, as determined by parental report. These two levels of age of acquisition were chosen so that the children in the sample should understand them in this context (to demonstrate receptive understanding given two visual alternatives), but that some words would be more recently acquired, to help maintain interest in the task (as recommended by Fernald, Perfors, and Marchman (2006)). In the original version of this task, the earlier developing words were presented in the first 20 trials and the later developing words in the second 20 trials.

Early trouble shooting with the task

After the first 10 children of the study, it was noted the children were not attending to the whole task, meaning few usable trials were available for calculating mean latency of spoken word recognition. The mean last trial attended to for these children was trial 31 (*SD* = 9.5). Several factors were noted as influencing the children’s attention from watching the recordings. Firstly, it was noted the children were getting frustrated with the tag questions, as they were answering both the initial question and the tag question, thus answering each item twice. The second possible factor was that there was a lot of repetition of the same pictures, as the earlier developing set of pictures was presented three to four times each over the first 20 slides. Therefore two changes were made. Firstly the tag questions were removed. On the first trial the training phrase: “point to it” was retained. Secondly the earlier and later developing stimuli were alternated throughout the slide show to reduce the immediate repetition. This was trialled with the next seven children. Due to an oversight when reorganising the slideshow, however, it was later discovered that the target picture now appeared on the left side for seven trials in a row (Trials 27-33). This was unfortunate as the children seemed to start anticipating this pattern and losing interest in the task during this series of trials. However despite this, a marked improvement in the length of time they attended to the second version of the task compared with the original was noted. The mean last trial attended to was now Trial 38 (*SD* = 2). The difference in mean last trial attended to across the two version of the task was significant, t (7.6) = -2.22, p=.03, one-tailed. This format was therefore kept for the remainder of the study (a total of 70 children were tested using this new version). Due to the minor nature of the changes, it has also been assumed that the first and second versions of the slide show are comparable, so the mean latencies from the first 10 children were included in the analysis.

Coding procedures

After the task was administered, eye movements were coded from the video using Elan (Max Planck Institute for Psycholinguistics, 2012) software. Eye movements were coded for each frame (0.04 s) for the two second window after the onset of the target word. Four codes were used: *left*, *right*, *away* or *between*. *Left* referred to looking at the left picture, and *right* at the right picture, *between* was looking between the two and *away* was anywhere else. The first trial for each child was considered a training trial and was not included in the analysis. Any trials where the child was inattentive or was not responding as expected, were also discarded. These decisions were made blind to the distractor / target status of the pictures, so that the researcher was not influenced by the need to increase the number of analysed trials containing distractor to target shifts. More detailed coding procedures are in Appendix E.

Trials were designated as “distracter initial” or “target initial” depending on which picture the child was fixating on at target word onset. The latency to shift from the distracter to the target picture, on distracter initial trials was calculated for each child. A script in Praat (Boersma, 2001) was used to extract the latency calculations from the annotations in Elan. A mean latency was calculated based on all distractor initial shifts initiated within 240-1000 ms from target word onset using Microsoft Excel.

Upper and lower cut-offs

A range of cut-offs (e.g. 200-400 ms for the lower cut-off) has been used for children in a similar age range (Bailey and Plunkett (2002); Ballem and Plunkett (2005); Fernald et al. (2006)). The shortest possible time that a distracter-to-target shift should take at 24-30 months old is unknown (see Swingley, Pinto, and Fernald (1999)). This task involves multiple processing components which can all vary individually. Given the individual variation in speed, some loss of valid information and or noise in the data is unavoidable. A cut-off of 240 ms was expected to be the fastest response time possible in reaction to the stimulus at this age. This relatively low cut-off was chosen so as to include as many trials as possible for the mean latency measure. The upper cut-off chosen (1000 ms) was two standard deviations above the mean of all valid shifts for all participants. It was felt that shifts slower than this were not representative of the children's pure reaction times; they were likely affected by other factors. A comparison was made of different cut-offs (200-300 ms for the lower, and 1000-1800 ms for the upper) and they were not found to make a significant difference to the results. After these cut-offs were applied, the number of trials each latency measure was based on for each child was 2-15, with an average of 7.97 trials. Children who had one or less distractor to target useable shifts were entered as missing data. It was decided to include those with mean latencies based on only two to three shifts despite the paucity of information on these children's range of latencies. There were only three children in this situation. The alternative of entering missing data for these three was judged to be more problematic for the subsequent analyses than including these scores. Imputing these values was attempted using multiple imputations, however the distribution of the latencies of the whole sample were heavily skewed, whereas the imputed values were based on an assumption of a normal distribution. Therefore the values imputed were not likely to have occurred naturally. Therefore these children's actual mean based on two or three shifts was entered as their latency values.

Inter-rater reliability for this coding system was checked. For 20% of the participants, 25% of trials which contained more than one shift were re-coded by research assistant. Trials with more than one shift were selected, as the main opportunity for inter-rater disagreement is in coding frames involving a shift. The coding of frames was found to be accurate within one frame 99.5% of the time. Coding of frames involving a shift was accurate within one frame 95% of the time. This meets the standard of reliability recommended by Fernald et al. (2008).

8. The Test of Early Non-Word Repetition (TENR)

The Test of Early Non-Word Repetition (TENR) is a test of 16 non-words designed for children aged two to four years (Stokes & Klee, 2009a). It contains four trials each of one to four syllable level words. Syllables are made up of combinations of CV and CVC structures and consonants which are in the phonetic inventories of two year olds. The test was extended to include 4 five syllable non-words since the original study was published. See Appendix D for the score form which contains the list of non-words. Unlike in the original study, the stimuli were presented in a PowerPoint slide show. The child was asked to repeat some "funny names" so that images depicting aliens would appear on screen. Note that the non-words were presented by live voice in Stokes and Klee (2009a). Recorded voice was preferred in this study so that the non-words were standard across presentations. The recordings were made by a female Australian speaker. NZ realisations of the Australian accent were accepted as correct responses. All syllables were given equal stress. Each word was heard only once. Children's responses were transcribed online while the child watched the alien appear on screen as a reward. These online transcriptions were checked against audio and video recording for accuracy later. One point was given for each phoneme correct which was tallied across all items attempted for a *total score*. Minor distortions (e.g. dental

/s/) were scored as correct. Additions were not penalised as they did not indicate loss of phonological information. Transpositions were counted as incorrect. If a child did not have a phoneme in their phonetic repertoire, any substitution errors on this phoneme were scored as correct. This was determined by any instance of the phoneme’s use in the TPT or other items of the TENR. As the TENR contains few later developing phonemes, it was unusual that a child did not have all the phonemes tested in their phonetic repertoire, with the exception of /r/, /l/, /s/ and /ʃ/, which some children were still developing. These phonemes accounted for 11% of the total score on the TENR.

A difficulty in measuring non-word repetition in two year olds is that those with limited expressive language often refuse or are unable to imitate novel words on one presentation. In order to measure the full range of variability in non-word repetition skills in this study, we assigned scores of zero to those children who were not yet imitating words. This was determined by their lack of response to the practise and initial test items on the TENR; confirmed by parent report that they were not yet imitating words; the child’s CDI scores and anecdotal observations of the child’s lack of response to imitation opportunities throughout the assessment sessions (e.g. during the phonology test). If a child refused to attempt the TENR, but there was no clear evidence they were not yet imitating words, their TENR score was entered as missing data.

Partial attempts at the TENR were included in the analysis. Across all the assessments children were generally keen to comply with adult directions, but once items became challenging, they began to become non-responsive. Therefore the assumption was made that children who started the TENR but then stopped attempting words partway through the test, could not actually repeat the words which they refused to attempt. Therefore the total scores of those who did not complete the whole test were included along with those who attempted all 20 items.

Inter-rater reliability was determined by subsequent rescoring of the TENR from the audio / video recordings by a research assistant. This was done blind to the child's previous TENR score and language status. The criteria for inclusion in the reliability study were that firstly, at least one item of the TENR must have been attempted. Secondly, only participants with a good quality audio / video recording of all attempted items were included. A total of 39 TENR recordings met these criteria. A high correlation between the two raters was achieved, r (37) = .99, p < .001, two-tailed. Inter-rater transcription accuracy was not compared as the total score on the TENR was based on the judgement of correct / incorrect. Inter-rater agreement on phoneme accuracy was 81%. However, the listening conditions were not identical. The first rater scored both online face-to-face and using the recordings, whereas the second rater using the recordings alone. The second rater scored the children consistently lower than the first. The average score by the first rater was 60.56 and by the second, 58.44. The paired samples t-test was significant, t (38) = 3.21, p = .003, two-tailed. The total score differed by more than 5% between the two raters in 36% of the re-scorings. No paired total scores differed by more than 16%. There was a discrepancy of five points or less for 32/39 of the scores. These differences were likely due to the difference in the listening conditions. Sound productions were slightly more difficult to perceive in the recordings than live. These inter-rater results also highlight the difficulty of accurate phonemic transcription of two year old children with emergent phonology and phoneme production.

3.4 Procedures

Parents brought their children to the Child Language Centre for two 1.5 hour sessions approximately a week apart (range = 1 day to 4 weeks). The time delay between the CDI date and these assessment dates was always one month or less, as parents updated the CDI if the delay was greater than this.

Each session lasted from 70 to 90 minutes. Sessions were audio and video recorded. The video cameras were wall mounted and had pan / tilt motion and zoom capabilities. Beyerdynamic boundary microphones were used for the audio recordings. These were set into the ceiling in the centre of the clinic rooms.

In most cases, the test protocol for the first session was the KWM I task, ROWPVT-4, VRO, LWL I task, TENR and the language sample. The book was then given to the child after the first session, and the parent took the BRIEF-P and an additional parent concern questionnaire (PEDS) home to return completed at the next session. The second session usually followed this format, the PLS-4 AC, A not B task, TPT, PLS-4 EC and the Otoacoustic Emissions (OAE) screening. The parent was then given a $20 voucher to thank them for participating. If there were any concerns about their child's development, these were discussed at this point and clinical information and advice given.

Two year olds are a difficult age group to assess using behavioural measures due to their frequent changes in mood, attention and motivation. Therefore a range of strategies were used to reduce these challenges. In order to avoid fatigue effects, breaks, behavioural and social reinforcement were used to gain as much participation from each child as possible. The order of tasks was varied if the child was not motivated by a particular activity. Some children were not developmentally able to participate in certain tests (e.g. not yet imitating words for the TENR), and these test scores were entered as a zero. Those who did not comply despite having the necessary developmental level (as determined by parental report or other assessment information) were entered as missing data. Occasionally parents agreed to bring their children in for a third assessment visit, if it was believed an additional visit would complete the test protocol.

Chapter Four: Study 1 Results

4.1 Missing data analysis

4.2 Descriptive statistics for the assessment data

4.3 Research question 1

4.3.1 Bivariate associations between the variables

4.3.2 Effect of demographic variables on expressive vocabulary

4.3.3 Initial multivariate regression model for expressive vocabulary at 24-30 months

4.4 Research question 2

4.4.1 Comparison multivariate regression model for expressive vocabulary

4.5 Research question 3

4.5.1 Criteria for late talking

4.5.2 Comparison of group means

4.6 Research question 4

4.6.1 Comparison of group means in working memory while controlling for language

4.7 Chapter summary

4.1 Missing data analysis

This study has a relatively complete data set. Seventy-nine children were administered 11 assessments each, for a total of 869 possible data points. Eight hundred and forty seven (98%) of these were able to be entered. Percentages of missing data from the behavioural assessments are summarised in Table 4.1. The Parent Questionnaire I was completed by 100% of the parents and over 99% of the questions in it had been answered across all participants. The least complete variable from this questionnaire was birth weight, which was missing for 3% of the participants.

Table 4.1

Percentage of missing data by assessment at 24-30 months

Measure	N completed	% missing
CDI	79	0%
Parent Questionnaire I	79	0%
KWM I	79	0%
ROWPVT-4	79	0%
PLS-4 AC / EC	79	0%
TPT	77	3%
A not B task	72	9%
TENR	78	1%
VRO	76	4%
BRIEF-P	77	3%
LWL I Mean latency	72	9%

Note. CDI = MacArthur Bates-Communicative Development Inventory; KWM I = The Key Word Working Memory I task; ROWPVT-4 = The Receptive One Word Picture Vocabulary Test (Fourth Edition); PLS-4 AC / EC = The Preschool Language Scale (Fourth Edition) Auditory Comprehension / Expressive Communication; TPT = Toddler Phonology Test; TENR = Test of Early Non-Word Repetition; VRO = The Mullen Early Scales of Learning: Visual Reception Organisation; BRIEF-P = Behaviour Rating Inventory of Executive Function - Preschool Version; LWL I = Looking While Listening I Task.

There were some detectable patterns of missingness in the 11 assessments across the participants, but only 25% of missing entries were due to an assumed poor ability in that assessment. (Note these missing entries do not refer to the assignment of zero scores, but rather where data was not entered). Given that there was 9% missing data for two variables and that there was some missingness not at random, multiple imputation was trailed for the results of Study 1 (Schafer & Graham, 2002). However it made very little difference to the results compared with using pairwise deletions. Multiple imputation also did not predict values of the A not B and LWL I tasks or the BRIEF-P questionnaire well. These assessments have skewed distributions and were not well predicted by the other variables in the study. This meant the imputed values for these assessments were not very likely. These three assessments had the highest rates of missing data. Taking these points into consideration and given the low number of missing data points overall, it was decided pairwise deletions were an appropriate way to manage the missing data in this study.

4.2 Descriptive statistics for the assessment data

Table 4.2 summarises the mean, standard deviation, range of scores, sample size and measures of skew and kurtosis of the assessment data for the whole sample at 24-30 months.

The measures performed well to capture variation in development in this age range. There were no ceiling effects on any measure. Minor floor effects could be seen on the TENR as 12 children were not yet imitating words on hearing a single presentation. There were also 13 zero scores in the phonology measure as these children did not have enough expressive vocabulary to name (or imitate) words in the TPT. In addition, the BRIEF-P scores (Inhibition, Shift, Emotional Control, Working Memory, Plan / Organise, GEC) all showed floor effects as is typical of behaviour rating scales, as most children were not considered to have problem behaviours in these areas by their parents.

The distributions of each variable were considered using z tests for skew and kurtosis. They were normally distributed except for the BRIEF-P scores and the LWL I task mean latency. Skewed distributions are expected for both behavioural rating scales and reaction time assessments. For the correlation and regression analyses, scores on these two assessments were transformed using a natural log transformation. The values of skew and kurtosis for these variables did not exceed 1.96 (and therefore were not significantly different from a normal distribution at the $p < 0.05$ level), so are unlikely to have caused a problem in these parametric analyses. The only exception to this was log transformed Working Memory variable, which had a kurtosis of 2.91. Kurtosis alone does not have a large impact on correlation and regression analyses, so this was not a concern (Miles & Shevlin, 2001). Non-parametric tests (Mann-Whitney *u* tests) were used for these variables when comparing the means between groups.

Table 4.2

Descriptive statistics for the assessment data at 24-30 months

	Total Sample (N = 79)				
Variable	Mean (SD)	Range	N	Skew (SE)	Kurtosis (SE)
CDI (words produced)	328.92 (194.71)	2-645	79	-.27 (.27)	-1.19 (.54)
ROWPVT-4 (raw score)	32.03 (11.49)	0-59	79	-.51 (.27)	.58 (.54)
ROWPVT-4 (standard score)	108.23 (9.94)	74-126	77	-.63 (.27)	.75 (.54)
PLS-4 AC (raw score)	34.87 (6.80)	19-50	79	-.03 (.27)	-.35 (.54)
PLS-4 AC (standard score)	108.61 (19.66)	53-150	79	-.38 (.27)	.11 (.54)
PLS-4 EC (raw score)	37.52 (7.52)	22-50	79	-.25 (.27)	-.73 (.54)
PLS-4 EC (standard score)	112.32 (23.16)	70-150	79	-.02 (.27)	-.98 (.54)
PLS-4 (total raw score)	72.52 (13.86)	43-98	79	-.17 (.27)	-.63 (.54)
VRO (raw score)	32.59 (4.49)	21-44	76	.16 (.28)	.21 (.55)
VRO (standard score)	113.41 (14.82)	69-133	76	-.65 (.28)	-.12 (.55)
TPT-PCC	52.34 (30.14)	0-95	79	-.67 (.27)	-.83 (.54)
TENR (total score)	44.83 (36.43)	0-116	78	.29 (.27)	-1.17 (.54)
KWM I (total score)	20.76 (7.40)	0-29	79	-1.45 (.27)	-1.15 (.54)
A not B task (total score)	14.00 (8.89)	2-29	72	.45 (.29)	-1.46 (.56)
BRIEF-P:					
Inhibit (raw score)	19.22 (4.51)	16-34	77	1.86 (.27)	3.14 (.54)
Shift (raw score)	11.68 (2.67)	10-22	77	1.93 (.27)	3.42 (.54)
Emotional Control (raw score)	12.34 (3.21)	10-23	77	1.47 (.27)	1.44 (.54)
Working Memory (raw score)	19.62 (4.69)	17-37	77	2.25 (.27)	4.54 (.54)
Plan / Organise (raw score)	11.70 (2.85)	10-23	77	1.97 (.27)	3.49 (.54)
GEC (total raw score)	74.65 (16.02)	63-132	77	1.87 (.27)	3.01 (.54)
LWL I (mean latency) (ms)	0.43 (0.09)	0.25-0.75	72	1.29 (.28)	2.93 (.56)

Note. CDI = MacArthur Bates-Communicative Development Inventory; KWM I = The Key Word Working Memory I task; ROWPVT-4 = The Receptive One Word Picture Vocabulary Test (Fourth Edition); PLS-4 AC / EC = The Preschool Language Scale, (Fourth Edition) (PLS-4) Auditory Comprehension / Expressive Communication; TENR = Test of Early Non-Word Repetition; VRO = The Mullen Early Scales of Learning: Visual Reception Organisation; BRIEF-P = Behaviour Rating Inventory of Executive Function - Preschool Version; GEC = Global Executive Composite; LWL I = Looking While Listening Task I.

4.3 Research question 1

What are the bivariate and multivariate associations between the expressive vocabulary skills of 24-30 month olds and the following aspects of the working memory system: PSTM, processing speed, VWM, VSWM and EF?

Research hypothesis

Significant associations will be seen between all aspects of the working memory system and expressive vocabulary on a bivariate level. PSTM is expected to show the strongest association with expressive vocabulary in the multivariate model.

4.3.1 Bivariate associations between the variables

Table 4.3 summarises the Pearson's bivariate partial correlations between measures for the whole sample. The zero-order correlations and the partial correlations (controlling for age) are included. One-tailed significance testing was used as a higher skill level in every assessment was expected to positively affect all other behavioural measures in the protocol.

These results were consistent with the hypothesis, with the exception of the A not B task. The TENR and KWM I tasks correlated with each other and all the language measures to a strong degree (r (75-76) > .5, p < .001 for all comparisons). The LWL I mean latency correlated to a moderate degree with the language measures. It had the strongest correlation with the non-verbal cognition measure (r (67) = -.60, p < .001). Of the six EF measures from the BRIEF-P, Shift, Emotional Control, Working Memory and GEC had significant small to moderate associations with expressive vocabulary. Contrary to expectations, the A not B task did not correlate to a significant degree with any measure other than Shift and then only with a small effect size (r (67) = .21, p = .04). This result was unexpected and may cast doubt on the suitability of this task for measuring VSWM in this age range. At the very least,

Table 4.3 Bivariate correlations between the variables at 24-30 months (N = 79)

Variable	1	2	3	4	5	6	7	8	9	10	**11**	**12**	**13**	**14**	**15**	**16**	17
1. CDI	_	.64***	.02	.81***	.75***	.76***	.90***	.86***	.81***	.39***	-.12	-.38***	-.25*	-.29**	-.16†	-.28**	-.32*
2. KWM I total score	.62***	_	.13	.52***	.76***	.73***	.71***	.72***	.62***	.62***	-.01	-.30**	-.10	-.14	.00	-.12	-.45***
3. A Not B total score	-.05	.09	_	.05	.14	.09	.10	.10	.12	.17†	-.05	.19†	-.02	.00	.05	.01	-.19†
4. TENR total score	.80***	.50***	.01	_	.74***	.68***	.82***	.77***	.79***	.29**	-.08	-.23*	-.18†	-.17†	-.04	-.17†	-.28**
5. ROWPVT-4 raw score	.73***	.74***	.06	.73***	_	.81***	.83***	.84***	.71***	.60***	-.16†	-.33**	-.27*	-.28**	-.16†	-.28**	-.39***
6. PLS-4 AC raw score	.75***	.73***	.01	.66***	.81***	_	.88***	.96***	.66***	.62***	-.16†	-.36**	-.26*	-.31**	-.13	-.29**	-.33**
7. PLS-4 EC raw score	.89***	.70***	.02	.81***	.82***	.87***	_	.97***	.83***	.55***	-.11	-.36**	-.25*	-.27**	-.14	-.26*	-.38***
8. PLS-4 total raw score	.85***	.71***	.01	.76***	.84***	.96***	.97***	_	.77***	.60***	-.15	-.38***	-.27**	-.30**	-.15	-.29**	-.36**
9. TPT-PCC	.79***	.61***	.04	.78***	.69***	.64***	.81***	.76***	-	-.40***	-.12	-.31**	-.28**	-.24*	-.13	-.23*	-.38**
10. VRO raw score	.32**	.61***	.03	.23*	.60***	.60***	.51***	.56***	.34**	_	.20*	.24*	.85***	-.28**	-.22*	-.25*	-.62***
11. Inhibit (log)	-.11	.00	-.02	-.06	-.17†	-.15†	-.11	-.14	-.11	-.20*	_	.56***	.74***	.85***	.87***	.92***	.18†
12. Shift (log)	-.39***	-.30**	.21*	-.23*	-.34**	-.37**	-.36**	-.39***	-.31**	-.25*	.56***	_	.69***	.63***	.64***	.77***	.22*
13. Emotional Cont. (log)	-.25*	-.10	.01	-.17†	-.27**	-.27*	-.25*	-.27**	-.20*	-.25*	.74***	.69***	_	.69***	.70***	.86***	.18†
14. Working Memory(log)	-.27*	-.12	.04	-.15	-.26*	-.29**	-.25**	-.29**	-.22*	-.26**	.85***	.64***	.69***	_	.91***	.93***	.21*
15. Plan / Organise(log)	-.14	.02	.08	-.02	-.15†	-.12	-.17†	-.14†	-.11	-.21*	.87***	.64***	.70***	.91***	_	.93***	.19†
16. GEC raw score (log)	-.26*	-.10	.05	-.15	-.27**	-.28**	-.33**	-.28**	-.22*	-.29*	.92***	.77***	.86***	.93***	.93***	_	.19
17. LWL I mean latency (log)	-.29**	-.43***	-.17†	-.26*	-.35**	-.30**	-.35**	-.33**	-.35**	-.60***	.16†	.21*	.17†	.20†	.18†	.18†	_

Note. Zero-order correlations are presented above the diagonal and partial correlations (controlling for age) are presented below the diagonal. Items in bold (11-16) are from the BRIEF-P assessment. CDI = MacArthur Bates - Communicative Development Inventory; KWM I = The Key Word Working Memory I task; TENR = Test of Early Non-Word Repetition; ROWPVT-4 = The Receptive One Word Picture Vocabulary Test (Fourth Edition); PLS-4 AC / EC = The Preschool Language Scale, (Fourth Edition) (PLS-4) Auditory Comprehension / Expressive Communication; TPT - PCC = Toddler Phonology Test – Percent Consonants Correct; VRO = The Mullen Early Scales of Learning: Visual Reception Organisation; GEC = Global Executive Composite; LWL I = Looking While Listening I Task. †p<.10. * p<.05. **p<.01. ***p<.001.

associations with the other measures which load on the CE (KWM I and the EFs) could have been expected, however only processing speed showed a small correlation and even that was not significant. Different interpretations of this result will be considered further in the discussion chapter.

4.3.2 Effect of demographic variables on expressive vocabulary

The research questions focus on the relationship between psycholinguistic processing skills and language; however there are well-documented effects of demographic variables on language also. These needed to be accounted for to allow a clearer picture of the unique variance that each working memory variable contributed to language variation. The sample size was too small to include all such variables simultaneously in the multivariate regression models, therefore I first identified which demographic variables had a significant relationship with expressive vocabulary (CDI) scores, so that only these would be included in subsequent analyses.

Firstly, a series of univariate analyses of covariance (ANCOVA) were run, which estimated the effect of the categorical demographic variables on CDI scores. Secondly, partial correlations were calculated for the interval data. An alpha level of .05 was set for all the significance testing in this study. Age was the control variable. Age correlated with the CDI scores to a moderate degree, as expected within this small age range, r (77) = .27, p =.008, one-tailed. The assumption of the homogeneity of variance was met in all the ANCOVA analyses. When this assumption is met, ANCOVA maintains excellent control of Type I errors (Rheinheimer & Penfield, 2001).

Boys had significantly lower scores on the CDI when controlling for age (M = 283.81, SE = 24.21) than girls (M = 426.36, SD = 35.65), $F(1,76)$ = 10.89, p = .001, partial eta squared = .13. The mean difference between groups was 142.55 words, 95% CI [56.52, 228.57].

Children with a positive reported family history for speech, language or learning difficulties in first- or second-degree relatives narrowly missed having a significantly lower mean on the CDI (M = 284.98, SE = 32.01) than those who did not (M = 362.12, SE = 27.80), $F(1,76) = 3.29$, $p = .07$, partial eta squared = .04.

While first-born children had a higher mean on the CDI (M = 359.37, SE = 27.90) than second- (M = 282.55, SE = 37.98), third- (M = 299.89, SE = 84.49) and fourth-born children (M = 296.93, SE = 109.08), the differences between the groups were not significant, F (3,78) = .96, $p = .42$, partial eta squared = .04. Pairwise comparisons between these groups likewise were not significant.

The effect of the number of children in the family on expressive vocabulary was evaluated. Children were split into five groups according to family size (the range was 1-5 children in each family). Note there was only one child with four children in the family and two with five children. There were significant differences in the means between groups, F (4,73) = 3.28, $p = .02$, partial eta squared = .15. Pairwise comparisons between groups showed that rather than there being a significant difference between every group, the significant differences were between having two, three or four children in the family, compared with one child. There was no difference between having one or five children (this could have been due to small number of families with five children). Therefore the data was regrouped according to "only-child" status. Only-children had a significantly higher CDI score (M = 426.27, SE = 36.17) than children-with-siblings (M = 283.86, SE = 24.41), F (1,76) = 10.46, $p = .002$, partial eta squared = .12. The mean difference between groups was 142.42, 95% CI [54.69, 230.14].

Children were grouped by their parent's highest qualification level. Six groups were formed based on the New Zealand Qualifications Authority (NZQA) framework (as shown in Table 3.1). Comparisons between these groups on CDI scores did not show significant differences, F (5,71) = 1.93, p = .10, partial eta squared = .12. Therefore parents were divided by "degree status" instead. This analysis showed children whose parents had degrees had significantly higher CDI scores (M = 373.31, SE = 25.35) than those who did not (M = 264.68, SE = 32.96), F(1,75) = 6.81, p = .01, partial eta squared = .08. The mean difference between groups was 108.63 words (95% CI [25.75, 191.50].

Bivariate partial correlations with the CDI scores were calculated, controlling for age. Hours per week in day care, r (76) = .03, p = .83, two-tailed; prematurity (weeks), r (76) = .01, p = .48, one-tailed, and birth weight, r (74) = .13, p = .13, one-tailed, were not significantly associated with the CDI scores.

Summary

As has been previously established in this age range, age and sex had a significant effect on expressive vocabulary (Dale & Fenson, 1996). Children whose parents had degrees had significantly higher expressive vocabulary scores than children whose parents did not, as expected from previous studies (Hart & Risley, 1995). More fine grained comparisons of parental education were not significant across all paired groupings. In this sample, children-without-siblings had a significantly higher expressive vocabulary than children-with-siblings as was found by Zubrick et al. (2007). Children with a positive family history had lower expressive vocabulary scores than children who did not, although this difference narrowly missed being statistically significant in this study, despite being found to be a significant factor in other studies in this age range (Reilly et al., 2007; Zubrick et al., 2007). Prematurity and birth weight were not associated with expressive vocabulary in this study. (Reilly et al., 2007) also reported that these variables were not associated with language development in this age range. However Zubrick et al. (2007) found that a proportion of optimal birth weight of less than 85% and gestation of less than 36 weeks were both significant predictors in their study. Hours in day care were also not associated with expressive vocabulary in this study, as has been previously reported (Zubrick et al., 2007). The demographic variables which significantly affected expressive vocabulary scores (child's age, sex, only-child status and parental education) will be included in subsequent regression analyses along with relevant behavioural measures.

4.3.3 Initial multivariate regression model for expressive vocabulary at 24-30 months

Having established the significant associations between the demographic variables and behavioural assessments with expressive vocabulary on a bivariate level, the next step was to build a regression model with multiple predictors. This is to establish the degree to which each variable contributed unique, rather than overlapping variance. The best model using demographic and behavioural variables for predicting expressive vocabulary was established using a mixture of hierarchical and backwards linear regression.

Age was first entered into the model. This removed the variance associated with age from the CDI scores. The next block followed a backward regression analysis procedure using variables which had significantly correlated with the CDI scores (KWM I, TENR, ROWPVT, PCC, VRO, Shift (log), Emotional Control (log), Working Memory (log) and LWL I mean latency (log)) and the demographic variables which had a significant impact on CDI scores (sex, only-child status, degree status). These were entered into the model simultaneously. One by one, starting from the highest p value, variables which were not significant ($p < .10$) were removed from the model. The model was re-estimated for the remaining predictors at each step. This continued until all variables were significant at the $p < .10$ level. Any predictors not significant at the $p < .05$ level at this point were then removed. Age was left in the model regardless of its p value relative to other predictors. The backwards method is preferred to the forwards method as it is less likely to generate suppressor effects (thus Type II errors) (Field, 2009).

The final model contained the following predictors: age, TENR, PCC, KWM I and Shift (log), sex and only-child status, and predicted 82% of the variance in CDI scores, *F* (7,67) = 43.03, $p < .001$. See Table 4.4 for a summary. The amount of unique variance age has with the CDI scores (words produced) (holding all other variables constant) was 2%; sex contributed 2% unique variance; Shift (log) 2%; only-child status 1%; PCC 1%, KWM I scores 1% and finally the TENR contributed 7% unique variance to CDI scores. These values were calculated by squaring the part correlations between each predictor and the outcome variable. There was a large amount of shared variance between the TENR and PCC, as if PCC was removed from the model, the unique variance contributed by the TENR increased to 22%. Amounts of unique variance from the other predictors remained similar.

Table 4.4

Multiple regression analysis summary for concurrent variables predicting expressive vocabulary (CDI scores) at 24-30 months (N = 77)

Model	B	SE β	Standardised β
(Constant)	169.30	198.692	
Age	13.26	5.47	.14*
KWM I	4.10	1.81	.16*
TENR	2.45	.47	.46***
PCC	1.31	.67	.20*
Shift (log)	-153.25	56.10	-.16**
Sex	-72.29	23.84	-.17**
Only child	-52.46	23.46	-.13*

Note. KWM I = Key Word Working Memory I Task; TENR = Test of Early Non-Word Repetition; PCC = Percentage Consonants Correct.
* p<.05. **p<.01. ***p<.001.

See Appendix F for the diagnostics procedure which was followed for this linear regression model.

Summary

The hypothesis was partially supported by this model. The variances had a substantial amount of shared variance. The TENR (PSTM) accounted for the most unique variance (7%)

in expressive vocabulary (CDI scores) at ages 24-30 months. This was despite the inclusion of PCC in the model, which reduced the amount of unique variance the TENR contributed from 22% to 7%. A novel finding here is the KWM I (VWM) and Shift also predicted significant unique variance. None of the other processing variables (VSWM, processing speed or other EFs) predicted significant unique variance in this multivariate model.

4.4 Research question 2

Do these aspects of working memory (PSTM, processing speed, VWM, VSWM and EF) improve on previously established multivariate linear regression models for predicting expressive vocabulary at 24-30 months?

Research hypothesis

Working memory measures will improve previously established multivariate linear regression models predicting expressive vocabulary at ages 24-30 months.

4.4.1 Comparison multivariate regression model for expressive vocabulary

Having established the best combination of concurrent predictors of expressive vocabulary at ages 24-30 months in this cohort, the next step was to determine whether the working memory variables explored by this study (EFs, speed, VWM, VSWM) improved on previously established models of expressive vocabulary in this age range. A forward hierarchical method was used. In this analysis, the order that variables were forced into the model was determined by the degree to which their influence on expressive vocabulary had been previously established in the literature. The main study for comparison here is Stokes and Klee (2009b) which reported a hierarchical multiple linear regression model containing age ($R^2\Delta$=.15), sex ($R^2\Delta$=.06), and the TENR ($R^2\Delta$=.24) to predict expressive vocabulary in a sample of 232 24-30 month old children. Only-child status has also previously been established as a predictor (Zubrick et al., 2007), as had phonology (Thal, Oroz & McCaw, 1995). In order to test the added value of working memory predictors, these were entered last. Therefore, age was forced into the model first followed by sex, only-child status, PCC and then the TENR. The Shift (log) and KWM I variables were entered in two final steps as the novel predictors. They were entered in this order, as Shift accounted for slightly more unique variance in the model using the enter method than the KWM I task. Table 4.5 shows the results of this series of analyses.

Table 4.5

Hierarchical multiple regression analysis summary for concurrent variables predicting expressive vocabulary (CDI scores) at 24-30 months (N = 77)

Step and predictor variable	β	SE β	Standardised β	R^2	ΔR^2
Block 1				.07	.07
Constant	-356.64	287.18			
Age	25.75	10.76	.27*		
Block 2				.19	.12
Constant	-346.95	270.46			
Age	29.05	10.18	.30**		
Sex	-142.55	44.38	-.34**		
Block 3				.31	.12
Constant	-431.95	253.35			
Age	36.02	9.71	.38***		
Sex	-144.86	41.38	-.35**		
Only-child	-144.81	42.09	-.35**		
Block 4				.70	.40
Constant	-178.60	169.30			
Age	12.49	6.86	.13†		
Sex	-48.38	29.09	-.12		
Only-child	-69.82	28.85	-.17*		
PCC	4.77	.49	.72***		
Block 5				.78	.08
Constant	-.163.55	147.41			
Age	12.81	5.97	.13*		
Sex	-65.87	25.58	-.16*		
Only-child	-55.21	25.30	-.13*		
PCC	2.32	.66	.35**		
TENR	2.45	.51	.46***		

Step and predictor variable	β	SE β	Standardised β	R^2	ΔR^2
Block 6				.80	.03
Constant	273.72	199.03			
Age	13.31	5.64	.14*		
Sex	-77.47	24.44	-.19**		
Only-child	-46.95	24.03	-.11†		
PCC	1.84	.65	.28**		
TENR	2.52	.48	.47***		
Shift (log)	-174.93	56.93	-.18**		
Block 7				.82	.01
Constant	169.30	198.69			
Age	13.26	5.47	.14*		
Sex	-72.29	23.84	-.17**		
Only-child	-52.46	23.46	-.13*		
PCC	1.31	.67	.20*		
TENR	2.45	.47	.46***		
Shift (log)	-153.25	56.10	-.16**		
KWM I	4.10	1.81	.16*		

Note. KWM I = Key Word Working Memory I Task; TENR = Test of Early Non-Word Repetition, PCC = Percentage Consonants Correct.
†p<.10. * p<.05. **p<.01. ***p<.001.

The final model (Block 7) was significant, $F(7,67) = 43.03$, $p < .001$. The predictors accounted for 82% of the variance as in the previous model. All the predictors in the model remained significant at the $p < .05$ level in the final block. The r squared change values were as follows: Age 7%; sex 12%; only-child status 12%; PCC 40%, TENR 8%; Shift (log) 3% and KWM I 1%.

Summary

This confirms previous research, that age and sex, non-word repetition and phonology contribute unique variance in expressive vocabulary at ages 24-30 months (Stokes & Klee, 2009b). Non-word repetition accounted for a smaller percentage of unique variance in expressive vocabulary scores (8%) than in the Stokes and Klee (2009b) study (24%), likely because of the additional of only-child status and PCC before non-word repetition in the current study. Only-child status added 12% significant unique variance once age and sex were entered. If the order of entry of PCC and the TENR were reversed, the relative beta weights also approximately reversed indicating a large amount of shared variance between these measures. A novel finding here is that the Shift and KWM I task and also added unique significant variance above that already accounted for (3% and 1% respectively). This finding supports the overall hypothesis that working memory variables will add predictive capability to previously established models of expressive language. This result is encouraging, representing a group level association between expressive vocabulary and aspects of concurrent working memory skills.

4.5 Research question 3

What differences in group means can be seen between the TD and late talking groups in the following aspects of working memory: PSTM, processing speed, VWM, VSWM and EF?

Research hypothesis

Performance on measures of working memory will be lower in late talkers than in TD children, particularly in PSTM.

4.5.1 Criteria for late talking

The sample was split into TD and late talker groups. As there is a range of criteria for what defines "late talking" in the literature, different options were carefully considered. Dale et al. (2003) reported that in their sample of two-year-old twins, children with a more severe delay initially (5^{th} percentile) were not at increased risk of ongoing difficulties compared with those below the 10^{th} percentile. Dale et al. (2003) also reported that initially TD children who fell into the disordered range by age four years had typically been just above the 10^{th} percentile at age two years. Therefore this evidence suggests the group of children one standard deviation below the mean or more (16^{th} percentile) were the ones best targeted for monitoring. In the current cohort, this criterion included all but one of the children who were not yet combining words together. Combining words is an important milestone linguistically. Support for the predictive importance of no two word combinations for weaknesses in oral language and general learning in middle childhood have been found by Poll and Miller (2013). Therefore the criteria for late talking in this study were a score of less than one standard deviation below the mean for expressive vocabulary and or no two word combinations. This meant there were 24 children classified as late talkers and 55 classified as TD.

4.5.2 Comparison of group means

Table 4.6 summarises the mean, standard deviation and range of scores of the behavioural measures for the TD and late talker groups. This was done to describe and compare the groups, and to investigate the likelihood that aspects of working memory play a role in determining late talker status. The two groups were not significantly different in age, t (77) = .82, p = .41, two-tailed, so age was not used as a control variable. *T*-tests were used to compare the assessments with a normal distribution. Group mean scores on the EF measures from the BRIEF-P and LWL I mean latency scores were compared using Mann-Whitney U tests, as their distributions departed significantly from normal (see whole group results for skew / kurtosis values). Cohen's d values were used as a measure of effect size. D values over 0.8 indicate a large effect, those over 0.5 show a moderate effect and over 0.2 indicates a small effect (Cohen, 1988).

The late talker group by definition was significantly lower than the TD group in expressive vocabulary (CDI words produced) and general expressive language skills (PLS-4 EC), with very large effect sizes (d = 3.59 for the CDI and d = 2.54 for the PLS-4 EC). Group differences on these variables, the TENR and the TPT, were larger than for any other variable measured. The late talkers scored significantly worse at receptive vocabulary (ROWPVT-4), general receptive language (PLS-4 AC) and VWM (KWM I); all showing large effect sizes.

This indicates that there were weaknesses in the overall linguistic system of the late talkers; their delays in expressive vocabulary are not likely to be just due to poor motor-speech planning and programming. The mean visual cognition score (VRO) was also lower for the late talker group with a moderate effect size. This may indicate there is some domain general weaknesses in the late talker group, or may it may be that better language skills contribute to performance on this measure. There was no evidence of differences between the group means in the A not B task. The LWL I mean latency measure was significantly slower for the late talkers, with a moderate effect size. The BRIEF-P scores showed a trend towards being poorer for the late talkers. The GEC (measure of overall EF) narrowly missed being significantly different between groups (p = .06), with a small effect size. The main differences in GEC scores arose from the Shift and Working Memory subscales. The late talker group was rated with significantly more problem behaviours in Shift (moderate effect size) and Working Memory (small effect size).

Table 4.6 Comparison of late talker and TD group results for the behavioural assessments

	Whole sample N = 79			Typically developing N = 55			Late talkers N = 24			Statistical tests for differences of means		
Measures	N	Mean (SD)	Range	N	Mean (SD)	Range	N	Mean (SD)	Range	t-test	d	Mann-Whitney
CDI (words produced)	79	328.92 (194.71)	2-645	55	436.13 (118.14)	164-645	24	83.25 (73.11)	2-292	16.17***	3.59	
ROWPVT-4 raw score	79	32.03 (11.49)	0-59	55	36.58 (8.47)	16-59	24	21.58 (10.76)	0-40	6.65***	1.54	
PLS-4 – AC raw score	79	34.87 (6.80)	19-50	55	37.55 (5.73)	25-50	24	28.75 (5.73)	19-35	6.55***	1.54	
PLS-4 – EC raw score	79	37.52 (7.52)	22-50	55	41.25 (5.10)	30-50	24	28.96 (4.55)	22-37	10.16***	2.54	
PLS-4 – total raw score	79	72.52 (13.86)	43-98	55	78.98 (12.25)	59-98	24	57.71 (8.93)	43-70	8.85***	1.98	
TPT - PCC	79	52.34 (30.14)	0-95	55	66.80 (18.67)	0-95	24	19.21 (24.83)	0-71	8.41***	2.17	
VRO – raw score	76	32.59 (4.49)	21-44	55	33.38 (4.34)	24-44	21	30.52 (4.31)	21-41	2.57**	0.66	
TENR – total score	78	44.83 (36.43)	0-116	54	62.24 (29.90)	0-116	24	5.67 (8.35)	0-31	12.82***	2.58	
KWM I – total score	79	20.76 (7.40)	0-29	55	23.15 (4.73)	5-29	24	15.29 (9.36)	0-27	3.89***	0.93	

	Whole sample N = 79			Typically developing N = 55			Late talkers N = 24			Statistical tests for differences of means		
Measures	N	Mean (SD)	Range	N	Mean (SD)	Range	N	Mean (SD)	Range	t-test	d	Mann-Whitney
A Not B I – total score	72	14.00 (8.89)	2-29	52	13.71 (8.63)	4-29	20	14.75 (9.86)	2-29	-.44	-0.11	
BRIEF-P :												
Inhibit (raw score)	77	19.22 (4.51)	16-34	54	19.20 (4.45)	16-34	23	19.26 (4.75)	16-33		-0.01	604.50
Shift (raw score)	77	11.68 (2.67)	10-22	54	11.07 (1.91)	10-19	23	13.09 (3.59)	10-22		-0.70	449.00*
Emotional Control (raw score)	77	12.34 (3.21)	10-23	54	11.89 (2.63)	10-20	23	13.39 (4.17)	10-23		-0.52	511.00†
Working Memory (raw score)	77	19.62 (4.69)	17-37	54	19.17 (4.35)	17-36	23	20.70 (5.36)	17-37		-0.31	480.50*
Plan / Organise (raw score)	77	11.70 (2.85)	10-23	54	11.56 (2.84)	10-23	23	12.04 (2.88)	10-19		-0.17	539.50
GEC (total raw score)	77	74.65 (16.02)	63-132	54	73.02 (15.05)	63-132	23	78.48 (17.86)	63-120		-0.33	479.00†
LWL I task (mean latency) (ms)	72	0.43 (0.09)	0.25-0.74	50	0.41 (0.06)	0.25-0.56	22	0.47 (0.12)	0.32-0.75		-0.65	389.50*

Note. CDI = Communicative Development Inventory; ROWPVT-4 = The Receptive One Word Picture Vocabulary Test (Fourth Edition); PLS-4 AC / EC = The Preschool Language Scale, (Fourth Edition) (PLS-4) Auditory Comprehension / Expressive Communication; VRO = Visual Reception Organisation; TENR = Test of Early Non-Word Repetition; KWM I = The Key Word Working Memory I task; BRIEF-P = Behaviour Rating Inventory of Executive Function - Preschool Version; GEC = Global Executive Composite; LWL I = Looking While Listening I Task. †p<.10. * p<.05. **p<.01. ***p<.001. All significance tests were one tailed.

4.6 Research question 4

Do any significant differences in working memory skills (PSTM, processing speed, VWM, VSWM and EF) remain when language differences between the two groups are controlled?

Research hypothesis

These effects are expected to be seen even when controlling for group differences in language skills.

4.6.1 Comparison of group means in working memory while controlling for language

The first overarching hypothesis for this research is that deficits in aspects of the working memory system are implicated in late talking, as predicted by the capacity limits theory of LI. If this is true, significant between-group differences in working memory should remain when controlling for group differences in language. If it is not, this would indicate support for the view that language and working memory skills develop as a unified system. A series of univariate analyses of covariance (ANCOVA) were run to test this hypothesis. The receptive language measure(s) most relevant to the particular working memory construct under investigation were used as controls. Expressive vocabulary was not used as a covariate, as these scores defined the groups. As working memory is hypothesised to contribute to variation in language and vice versa, it is expected that controlling for language will also remove some of the variance associated with the working memory constructs themselves. Therefore these ANCOVAs, if significant, would add strong support for the concept of late talking being influenced by capacity limitations in working memory processing. Null findings however, would not rule out the possibility that these aspects of working memory may constrain language development. The results are as follows:

When receptive vocabulary was controlled, late talkers still had significantly lower scores on the TENR (M = 21.38, SE = 5.18) than TD children (M = 55.26, SE = 3.17), $F(1,75) = 26.21$, $p < .001$, partial eta squared = .26. The estimated difference between groups on the TENR was 33.88 points, 95% CI [20.69, 47.06]. Given that the late talkers had lower scores in expressive phonology than the TD group, and that this could affect performance on the TENR, TPT-PCC was added as a second covariate in an additional ANCOVA. The group differences between late talkers (M = 31.40, SE = 5.47) and TD children (M = 50.80, SE = 3.15) remained significant when both receptive vocabulary and phonology scores were controlled, $F(1,74) = 7.23$, $p = .009$, partial eta squared = .09. The estimated difference between groups on the TENR was 19.40 points, 95% CI [5.03, 33.77].

Late talkers no longer had significantly lower scores on the KWM I task (M = 19.83, SE = 1.22) than TD children (M = 21.17, SE = .74) once receptive language scores (PLS-4 AC raw scores) were controlled, $F(1,76) = .74$, $p = .39$, partial eta squared = .01.

Late talkers no longer had significantly slower processing speeds on the LWL I task (M = .45 s, SE = .02) than TD children (M = .42 s, SE = .01), once receptive vocabulary scores (ROWPVT-4 raw scores) were controlled, $F(1,69) = 1.25$, $p = .27$, partial eta squared = .02.

The assumption of homogeneity of variance was not met for the ANCOVAs comparing the TENR, KWM I and LWL I group mean scores. However the assumption of normality of residuals held for these analyses, and a linear model was an appropriate fit for these data. ANCOVA has been shown to be a robust analysis when only one of these assumptions is violated (Rheinheimer & Penfield, 2001). This was not the case for the ANCOVA models comparing Shift and Working Memory scores across the two groups. The high number of zero scores on each of these EF subscales resulted in heavily skewed distributions and linear modelling was not appropriate for these data. Transforming the data did not improve model fit. Non-parametric options to compare the effect of receptive language ability on groups differences in Shift and Working Memory scores were explored (e.g. removing participants with zero scores), but no suitable solutions were available which retained all the data points. Therefore these two variables (Shift and Working Memory) were not examined in relation to Research Question 4.

Summary

Overall, partial support of the hypothesis was found. Aside from expressive language itself, the TENR (measuring PSTM) was the variable which most differentiated the late talking and TD groups. Large group differences were seen on the KWM I task (VWM); a moderate difference in the LWL mean latency scores (processing speed) and there was a small increase in parent reported problem behaviours associated with poor Shift and Working Memory in the late talker group. There were no significant group differences in the A not B task, Inhibit, Emotional Control, Plan / Organise and total GEC scores. However, once language was controlled, the only working memory measure which remained significantly different between the TD and late talking groups was the TENR (PSTM). Group differences in receptive vocabulary and phonology did not account for the poor performance on the TENR seen in the late talker group. Capacity limitations in PSTM may be constraining early expressive vocabulary growth in the late talker group. Group differences in the other working memory variables (VWM and processing speed) were no greater than expected, given the late talkers' weaker receptive language skills. Due to difficulties with the distribution of scores in the BRIEF-P measures, Shift and Working Memory were not able to be examined in relation to this question.

4.7 Chapter summary

In summary, the strongest predictor of expressive vocabulary on a group level was the TENR (PSTM), which accounted for 8% unique variance in the hierarchical model. If PCC was removed from this model however, this figure rose to 22%. This reflects the high degree of overlapping skills assessed by these two measures in this age range. The TENR remained significantly lower in the late talker group once the group differences in receptive vocabulary and phonology were controlled. This indicates late talkers may have a particular deficit in PSTM. Given the extensive body of research positing a role for PSTM in new word learning, these results suggest that early PSTM deficits may play a role in late talking for some children.

VWM was also uniquely associated with expressive vocabulary even when entered last in the hierarchical regression model, accounting for 1% unique variance. This is a novel finding which suggests a role for VWM in early vocabulary acquisition. This makes sense given that VWM seems to play a role in conscious processing of language for comprehension or production of new or complex language. A high percentage of language encountered must be new and or complex to two year olds, meaning a higher VWM span would assumedly be an advantage in acquiring comprehension and production skills. This was confirmed by an addition multiple linear regression analysis (not reported here) which confirmed 4% unique variance was contributed by the KWM I task in predicting concurrent receptive vocabulary. In terms of a group comparison however, the large group difference in KWM I scores was accounted for by the group differences in receptive language skills. This indicates that poor VWM skills are not a particular feature of late talkers, but seem to be associated with poor receptive language.

Shift was the only other aspect of working memory associated with expressive language in the regression models, predicting 3% unique variance in the hierarchical model. This is another novel finding, as Shift has not been previously significantly associated with early vocabulary acquisition. The likely direction of relationship between language and Shift will be considered further in the discussion when the results from both time points can be taken into account.

Processing speed, despite correlating with expressive vocabulary on a bivariate level, did not predict unique variance in multiple regression models. This may indicate that the variance associated with processing speed is subsumed within other measures such as the TENR and KWM I tasks, as speed is an important component of the working memory system. This may also apply to Inhibit, Emotional Control, Working Memory and Plan / Organise, which also did not contribute unique variance in the multiple regression models. There were also no group differences in processing speed over and above the differences expected by their lower receptive vocabulary skills.

Study 2 repeats this methodology after an 18 month period to examine the stability of relationships between variables over time. The RLTs at this point will be compared with the TD group to see whether aspects of working memory could account for their pattern of early delay followed by resolution of language skills. The fact that poor PSTM was characteristic of late talker status may indicate that other aspects of working memory may play a greater role in predicting later language outcomes, as it is expected that a significant number of the late talkers (who all have poor PSTM) will have typical expressive language when reassessed in Study 2. It is possible that those who have strengths in other aspects of the working memory systems at 24-30 months will have better outcomes 18 months later.

Chapter Five: Study 2 Methodology

5.1 Inviting participants to return

5.2 Descriptive statistics for demographic variables

5.3 Changes to the measures

5.4 Procedures

5.1 Inviting participants to return

Families were contacted twice during the 18 months between initial and outcome assessments. The first contact was to send the child a birthday card and the second was a newsletter updating them on the progress of the study. The newsletter was sent two months before the first participants were due to return for the outcome assessments. This was to keep families informed about how the study as a whole was going and to serve as a reminder. Target reassessment dates were set 18 months after the date on the child's CDI questionnaire. Parents were contacted two to three weeks out from this date to ask if they could bring their children back for the outcome assessments over the next month. At this point parents were reminded that they would be given a $20 shopping mall voucher, a $10-$20 petrol voucher (more was given to families living outside the city limits) and a book for their child for completing the study.

All families were successfully contacted. Only one family declined to participate in the outcome assessments as they had moved from the Christchurch region. Two other families had also moved from Christchurch but were able to bring their children in for assessments when visiting. These two families were offered $100 petrol vouchers and $20 shopping mall voucher in appreciation of their continued participation. Another two families were unable to attend at the target dates due to travel plans, but were able to come in a few months earlier / later to complete the assessments. Every effort was made to accommodate families' other commitments to minimise attrition. Financial assistance with childcare costs for younger siblings was offered to families who indicated a need. This offer was taken up by one family. Overall, 78 out of 79 children returned to complete the outcome assessments. Of these, all children completed both visits. Four participants returned for a third session so that more assessments could be completed.

The mean length of time between the first assessment appointment for the initial assessments and the follow up assessments was 18.24 months (*SD* = 1.0). The minimum time between visits was 16 months and the maximum was 23 months. Seventy-two of 78 children's time between assessments was 18 ±1 month. In total, 78 children were seen between the ages of 41-49 months, with a mean age of 44.82 (*SD* = 2.05) months. Two children's ages were outside the 42-48 month old target window at follow up.

5.2 Descriptive statistics for demographic variables

For the most part, the participants' demographics remained the same between Study 1 and Study 2. A summary of updates on demographic variables is provided in Table 5.1.

Table 5.1

Summary of child characteristics at ages 41-49 months

	Total Sample (N = 78)		
Variable	% (N)	M (SD)	Range
Age (months)		44.82 (2.05)	41-49
Number of children in the family:			
1	17 (13)		
2	58 (45)		
3	22 (17)		
4	2 (1)		
5	3 (2)		
No health problems including ear infections (reported)	47 (36)		
Other language spoken at home	10 (8)		
Hours per week in day care		20.03 (9.49)	0-45
Family history of speech language or learning difficulties	27 (21)		
Hearing screen (OAE screen at 2000, 2500, 3200 and 4000 hz at 50 dB):			
'Pass' one ear and 'refer' for the other	17 (13)		
'Pass' both ears	63 (49)		
'Refer' one ear and no data for the other	13 (10)		
Noncompliant (both ears)	8 (6)		

Note. OAE = Otoacoustic Emissions

Many families had another child between initial and outcome assessments. Only the number of families with four to five children remained constant. More children had reported health concerns over the last 18 months, with the percentage of children with no reported concerns dropping from 89% to 46%. Day care hours increased from 13.04 (*SD* = 12.27) hours per week to 20.03 (*SD* = 9.49). The number of families who reported speaking a second language at home dropped from 18% to 10%. Parent reported estimates of second language exposure ranged from 2% to 20-30%. While children with second language exposure of over 20% were excluded from Study 1, this estimate had changed for some families over the last 18 months, with two families now reporting they spoke a second language over 20% of the time. One of these children scored in the low range for receptive language using the norms on the PLS-4 at both initial and outcome assessments. This result should be interpreted with caution given his near bilingual status. Parent report of family history of speech language or learning difficulties for first and second degree relatives also changed from initial to outcome assessment. While the percentages of children with a reported family history was similar (Study 1 had 30% (*n* = 24) and Study 2 had 27% (*n* = 21)), the individuals with reported histories differed between the two time points. Therefore for the purposes of the regression analyses, a parent report of positive family history at either initial or outcome assessment was counted as a positive family history, on the assumption that this was more accurate than considering parent report at just one time point. Using this metric, 44% (N = 34) of children in the study had a reported positive family history. This figure was used in regression analyses in both Study 1 (ages 24-30 months) and Study 2 (41-49 months).

5.3 Changes to the measures

Changes were made to several of the measures between Studies 1 and 2. Changes were kept to a minimum to avoid introducing addition sources of variance in test performance. Firstly, a second parent questionnaire (Parent Questionnaire II) was developed to replace the one used in Study 1 (see Appendix G). Its main purpose was to update demographics, contact details and parent concern. It also repeated some questions asked in Study 1 to check for accuracy of parent report (e.g. family history). Secondly, the TPT was replaced with the Diagnostic Evaluation of Articulation and Phonology (DEAP) as this was the comparable measure by the same research group designed for this age range (Dodd, Hua, Crosbie, Holm, & Ozanne, 2002). Finally, while all the behavioural measures performed well in Study 1 to capture the range in performance, it was likely ceiling effects would be seen on the KWM, A not B and LWL tasks for 41-49 month old children. These three tasks were therefore extended, as outlined below. The extended versions were developed through a pilot study of 13 children aged 34-48 months old. The younger children were included to simulate delayed development.

1. Key Word Working Memory II Task (KWM II)

At 24-30 months of age, the KWM I task captured the variance in VWM development well. The mean score was 20.73 (*SD* = 7.39), with a range of 0-29. The maximum score possible on this task was 30. Some children were scoring near the ceiling on this task in Study 1 and it was expected they would have a longer VWM span 18 months later. Therefore the task was extended through a piloting process for the second study.

This new version had a maximum score of 63 key words correct (three trials each at 1-6KWLs). See Appendix D for the score form. In the pilot study, the average score was 52.38 (*SD* = 10.81) and the range was 22-60. The task ran smoothly and had face validity for measuring VWM in this age range. The toys array at this point was unchanged: pig, cat, dog, girl, bed, bath, drink, food. The two extension levels followed the same pattern of administration as the first four. As the task was now longer, different basal and ceiling criteria were needed to maximise efficiency in administration time. The basal was set as the lowest KWL at which the child achieved a score of 100%. The ceiling was reached when a child scored less than 50% of possible points at a KWL. The ceiling was set quite high as some children in the pilot study demonstrated inconsistent performance from item to item. It was noted the likelihood of children guessing correctly at 6KWL was quite high, as serial order was disregarded and there were only two extra toys present than needed for each instruction. To reduce the chance of correct guesses, two extra toys were added to the task (man and book). The number of extra toys available for each instruction was also standardised so that there were always four spare toys at each KWL. This meant that at the one key word level, five toys were on display and an extra toy was introduced at each KWL above this until all 10 toys were present for the 6KWL. Other than these changes, this task was administered in the same way as in Study 1.

2. A not B task (extended version)

This task was chosen as studies reported its ability to measure developmental change in executive functioning in the visual spatial domain in the age range 24-48 months (Epsy et al., 1999; Ewing-Cobbs, Prasad, Landry, Kramer, & DeLeon, 2004). The mean score in Study 1 was 14.00 (*SD* = 8.89) and the range was 2-29. This task had a maximum score of 30. As some children were scoring nearly at ceiling at in Study 1, an extension to the task was necessary to avoid ceiling effects with the 41-49 month old children. Garon et al. (2008) outlined five variables which could be manipulated to increase this task's ability to measure developmental change. Firstly, hiding locations which are more visually similar made the task more difficult. The original version used in Study 1 had two identical cups, so this was not a factor which could be manipulated. Secondly, closer proximity between the two hiding locations made it more difficult to distinguish which cup the reward was under. The original task had the cups 1-2 cm apart, so once again this could not be made more difficult. Thirdly, increasing the length of delay made the task more difficult. The original task tested at delay levels: 5, 10 and 15 s. Further increases in the delay time were discounted as an option as it was judged unlikely that 41-49 month old children would remain on task well with delays of over 15 s and this level of delay was the maximum reported in the literature.. Fourthly, some studies have increased the number of hiding locations to increase task difficulty. However Marcovitch and Zelazo’s (1999) meta-analysis of the A not B task discredited this practise as they found it actually decreased the chance of a perseverative error. Finally increasing the strength of the response set before switching was considered to increase task difficulty. In Study 1, the children needed to perform two consecutive correct searches before a reversal trial; this could be increased to three in Study 2. Unfortunately, this was the only change which could be made to the task to make it more difficult.

This adaptation to the task was trialled in the pilot study. The first seven children were administered the task with two consecutive correct responses needed before the reversal trial and the last six children were required to find it three times before switching. The results are outlined in Table 5.2.

Table 5.2

Pilot study results for the comparison of two versions of the A not B task

Task version	Mean age (months)	N	Mean (SD)	Range
Two consecutive correct	42.14	7	25.43 (5.47)	17-30
Three consecutive correct	41.50	6	24.67 (6.02)	16-29

The mean difference between these versions was slight and not statistically significant, t (11) =.24, p = .41, one-tailed. However the trend was in the expected direction and the sample size was small. Given that there were no other ways the task difficulty could be increased, and that there was evidence to indicate this should be effective (Garon et al., 2008), this change was adopted for Study 2. See Appendix D for the score form.

3. Looking While Listening II Task (LWL II task)

While it is not ideal to make major changes in methodology between sampling times, manually coding the video files frame by frame for Study 1 took several months of full time work. Repeating this methodology was not practical. Rather than omit the task from the test protocol, it was decided to use newly purchased eye tracking technology (Tobii X120) to run the LWL task in Study 2. The Tobii Eye Tracking Software (Tobii Technology, 2012) and E-Prime 2.0 (Psychology Software Tools Inc., 2010) with Tobii extensions (Psychology Software Tools Inc., 2012) were used to run the experiment. Microsoft Access was used to extract the required mean latencies from the data output. The experiment followed the same general procedure to measure speed of spoken word recognition as for Study 1, with a few improvements and extensions as outlined below.

A new set of words was chosen for the participants now that they were older. This was done in as similar a fashion as possible to Study 1. Half of the words chosen had an age of acquisition ranging from 36-41 months (mean age of acquisition of 38.5 months) and the other half from 42-48 months (mean age of acquisition of 44.5 months), based on (Morrison, Chappell, and Ellis (1997). The criteria for "acquired" was when 75% of children in their sample had this word in their expressive vocabulary. Some synonym replacements were made for the NZ context for example, "lorry" was changed to "truck" and "jigsaw" to "puzzle". This word list therefore had the same level of word difficulty relative to the children's ages as the word list in in Study 1.

There were several difficulties with the LWL I task in Study 1, so the opportunity was taken to mitigate these in Study 2. Firstly, the carrier phrase ("look at the") was recorded separately from the target word (c.f. the method used for the LWL I task as outlined in Appendix E: Determining the start of the word window). The audio files for the target words were cut to start directly on the onset of the first sound of the target word (for plosives, this started from the burst of the sound, not at the glottal stop) and cut after the word finished. This was necessary so that the timing of the onset of the target word could be tracked by the software. There was a slight loss of naturalness in the phrasing of the target phrase as a result, but this did not seem to affect the children's responses.

The second concern with the LWL data from Study 1 was that 15/75 children had fewer than five distractor-to-target shifts used calculate their mean latencies. Several strategies were applied to improve these statistics. As the children were now older, the number of trials was increased from 40 to 50. Boredom seemed to be a factor in not attending to trials in Study 1; therefore the amount of repetition of stimuli was reduced. In Study 1, there were 21 target words, so each corresponding picture stimulus was seen three to four times throughout the task. In Study 2, this was increased to 50 target words, so the children would see each one only twice, once as a target and once as a distractor.

In addition, using E-Prime and the eye tracker allowed greater flexibility in experimental design. In order to increase the number of trials where the child made a distractor to target shift, the target word audio clip was set to play once the child had been looking at the distractor picture for 200 ms. In other words, the stimulus presentation paused between the carrier phrase and target word until the child looked at the distractor. This change also meant if the child was distracted, the presentation automatically paused until they looked at the distractor again. Anecdotally, the children did not seem to notice this pattern and there was no evidence of them anticipating shifting to the other picture.

More reinforcement external to the task was added to motivate the children to continue watching until the end. A short film, "Ormie" (Silvestri, 2009), was segmented into 4 s clips which were played in a small window in the centre of the screen at the beginning of the task and after every second trial until end of the task. It was played in sequential order, with the main elements of the plot represented by the segments chosen. This acted not only as a motivator, but also gave the children a central fixation point to keep their attention towards the screen, and as the story unfolded, gave them a sense that the task was coming to an end.

The timing of the transitions between stimuli was reduced as the children were now older. Each movie clip lasted 5 s; each picture stimulus was shown for 1.6 s; a blank screen was shown for 1.5 s, then the next picture stimulus for 1.6 s, and the next movie clip ran for five s and so on. There were 25 of these cycles in the task. The whole experiment took approximately four minutes to run.

Other than these changes, the two versions of the tasks remained comparable. The picture choice and task design followed the same set of guidelines as the first version of the task.

The new version of the LWL task was trialled on the 12 children in the pilot study. Eleven of the children completed the task. All of these children watched through until the last trial. The mean number of shifts the children's mean latencies were based on was 21.3 with a range of 5-33. This showed great improvement from Study 1 where the mean was 7.97 and the range 0-15. Overall the adaptations and improvements to the task worked well with the older children. The mean latency for each child could be extracted from the eye tracker data output using Access within minutes of the task completion. Therefore the task was able to be included in the test protocol for Study 2.

5.4 Procedures

The session procedures were identical to the initial assessment session protocols in Study 1, with the following exceptions: the CDI was not completed; the KWM, A not B and LWL tasks had all been extended as outlined above; Parent Questionnaire II was used. As in Study 1, there were two assessments in the test protocol of the wider study which are not part of this dissertation (the language sample and a parent concern questionnaire (PEDS)). For nearly all the children, the first session ran in the following order, but it was changed if needed to maximise participation from the child: KWM II task, PLS-4 AC, TENR, ROWPVT-4, DEAP and the language sample. Children were given their book at this point. Parents were given the Parent Questionnaire II, PEDS and the BRIEF-P to complete between sessions. The second session usually followed this format: PLS-4 EC, LWL II task, VRO, A not B task (extended version) and the OAE screen. Parents received the gift vouchers at the end of this session. If there were any concerns about the children's development, these were discussed at the end of the second session and clinical information and advice was given.

Chapter Six: Study 2 Results

6.1 Missing data analysis

6.2 Descriptive statistics for the assessment data

6.3 Research question 5

- 6.3.1 Bivariate associations between the concurrent variables
- 6.3.2. Effect of demographic variables on expressive language at 41-49 months
- 6.3.3 Initial multivariate regression model for concurrent expressive language at 41-49 months

6.4 Research question 6

- 6.4.1 Comparison multivariate regression model for concurrent expressive language at 41-49 months

6.5 Research question 7

- 6.5.1 Comparison of group means at 41-49 months between RLT and TD groups

6.6 Research question 8

- 6.6.1 Comparison of group means in working memory while controlling for language at 41-49 months

6.7 Chapter summary

6.1 Missing data analysis

There were nine measures used in this study (the A not B task (extended version) having been removed from Study 2; see next section for an explanation). The percentage of missing data from each is summarised in Table 6.1.

Table 6.1

Number and percentage of missing data for each measure (N = 78)

Measure	Number completed	% missing
Parent Questionnaire II	76	3
KWM II	74	5
ROWPVT	77	1
PLS-4 AC	78	0
PLS-4 EC	77	1
TENR	75	4
VRO	76	3
DEAP	77	1
BRIEF-P	75	4
LWL II (Mean latency)	77	1

Note. KWM II = The Key Word Working Memory task, Version II; ROWPVT-4 = The Receptive One Word Picture Vocabulary Test (Fourth Edition); PLS-4 AC / EC = The Preschool Language Scale, (Fourth Edition) (PLS-4) Auditory Comprehension / Expressive Communication; TENR = Test of Non-Word Repetition; VRO = The Mullen Early Scales of Learning: Visual Reception Organisation; DEAP = Diagnostic Evaluation of Articulation and Phonology; BRIEF-P = Behaviour Rating Inventory of Executive Function - Preschool Version; LWL II = Looking While Listening Task, Version II.

A total of 97.7% (762/780) of the possible data points were present. Of the 3% (18) missing data points, 33% (6) were from one child (diagnosed as autistic after Study 1); parents failing to return questionnaires accounted for 28% (5) and the remaining 39% (7) were due to other children's non-compliance. Due to the overall low percentage of missing data, we decided that pairwise deletions were an appropriate way to manage the missing data.

6.2 Descriptive statistics for the assessment data

Table 6.2 summarises the mean, standard deviation, range of scores, sample size and measures of skew and kurtosis of the assessment data for the whole sample tested at ages 41-49 months.

Table 6.2

Descriptive statistics for the behavioural assessments at outcome (aged 41-49 months)

	Total Sample (N = 78)				
Variable	Mean (SD)	Range	N	Skew (SE)	Kurtosis (SE)
ROWPVT-4 (raw score)	59.68 (12.52)	29-91	77	.02 (.27)	-.084 (.54)
ROWPVT-4 (standard score)	112.13 (11.27)	85-141	77	-.04 (.27)	-.07 (.54)
PLS-4 AC (raw score)	51.41 (6.51)	21-62	78	-1.69 (.27)	5.51 (.54)
PLS-4 AC (standard score)	114.92 (16.04)	50-142	78	-1.01(.27)	2.57 (.53)
PLS-4 EC (raw score)	55.00 (7.21)	28-65	77	-1.60 (.27)	3.16 (.54)
PLS-4 EC (standard score)	117.84 (17.26)	50-148	77	-1.23(.27)	2.70 (.54)
PLS-4 (total raw score)	106.43 (13.46)	49-127	77	-1.65 (.27)	4.09 (.54)
PLS-4 (total standard score)	118.17 (17.61)	74-145	75	-1.13 (.27)	2.39 (.54)
VRO (raw score)	44.51 (3.34)	34-50	76	-1.11 (.28)	1.73 (.55)
VRO (standard score)	113.21 (15.45)	74-145	75	-.12 (.28)	.32 (.55)
TENR (total score)	97.61 (21.67)	18-127	75	-1.39 (.28)	2.23 (.55)
KWM II (total score)	48.34 (10.61)	19-60	74	-1.43 (.28)	1.10 (.55)
DEAP PCC	78.45 (14.92)	35-100	77	-.1.00 (.27)	.55 (.54)
A not B task (summary score)	3.61 (.84)	1-4	69	-2.02 (.29)	2.80 (.57)
BRIEF-P:					
Inhibit (raw score)	19.75 (4.42)	16-33	75	1.29 (.28)	.96 (.55)
Shift (raw score)	12.45 (3.04)	10-27	75	2.04 (2.8)	6.21(.55)
Emotional Control (raw score)	13.23 (3.23)	10-25	75	1.47 (.28)	2.27 (.55)
Working Memory (raw score)	20.03 (4.56)	17-36	75	2.00 (.28)	3.61 (.55)
Plan / Organise (raw score)	12.33 (3.10)	10-24	75	1.56 (.28)	2.23 (.55)
GEC (total raw score)	77.63 (15.40)	63-124	75	1.45 (.28)	.56 (.54)
LWL II task (mean latency) (ms)	.51 (.07)	.38-.68	77	.39 (.27)	-.12 (.54)

Note. ROWPVT-4 = The Receptive One Word Picture Vocabulary Test (Fourth Edition); PLS-4 AC / EC = The Preschool Language Scale, (Fourth Edition) (PLS-4) Auditory Comprehension / Expressive Communication; VRO = The Mullen Early Scales of Learning: Visual Reception Organisation; TENR = Test of Non-Word Repetition; KWM II = The Key Word Working Memory II task; DEAP - PCC = Diagnostic Evaluation of Articulation and Phonology Percentage of Consonants Correct; BRIEF-P = Behaviour Rating Inventory of Executive Function - Preschool Version; GEC = Global Executive Composite; LWL II = Looking While Listening II task .

The distributions of each variable were examined for skew and kurtosis using standardised scores. Most measures showed a positive skew towards the high end of achievement, but were still approximately normally distributed. The only scores with a high skew (z > 2) were the Shift and Working Memory scores from the BRIEF-P. All the BRIEF-P scores (except GEC and Inhibit) showed a high kurtosis, and some values were unusually high (for example Shift had a score of 6.21). I decided to use a natural log transformation on all the BRIEF-P scores, rather than to treat them in a piecemeal fashion. After the log transformation, all of the skew and kurtosis scores improved and none of the values were now over 1.6 for skew and 1.91 for kurtosis. This minor departure from normality should not impact on the results of the correlation and regression analyses (Miles & Shevlin, 2001). Non-parametric tests (Mann-Whitney *u* tests) were used to compare the means on the BRIEF-P. The LWL II task did not show a skew, which may indicate this task was not measuring reaction times as it did at age two years. This is considered further in the discussion chapter.

All the measures captured variance in scores in this developmental range well, except for the A not B task (extended version). Despite studies reporting that it is useful for capturing progress in this age range (Epsy et al., 1999; Ewing-Cobbs et al., 2004), in this sample little variance in performance was observed. Strong ceiling effects were noted in the first 10 participants. Although some errors were made, these appeared to be due to boredom rather than task difficulty (errors were more frequent in the second half of the task). Consequently, the task was shortened to capture the delay level at which an A not B error could be elicited, rather than gaining a score from 10 trials at the child's delay level. Testing began at the 15 s level. If children had no errors after four trials at this level (including one reversal trial after three consecutive correct searches) they were given a 'summary score' of four points. Children with an error at 15 s scored three points, provided they did not then err on four trials at the 10 s delay level. In the same way, children with an error at 10 seconds, but not 5 s, scored two points and children with an error only at 5 s delay scored one point. 55/69 children scored four points. This ceiling effect was so pronounced that the decision was made to remove this assessment from subsequent analyses.

6.3 Research question 5

What are the bivariate and multivariate associations between the expressive language skills of 41-49 month olds and the following aspects of the working memory system: PSTM, processing speed, VWM, VSWM and EF?

Research hypothesis

Significant associations will be seen between all aspects of the working memory system and expressive language on a bivariate level. PSTM is expected to show the strongest association with expressive vocabulary in the multivariate model.

6.3.1 Bivariate associations between the concurrent variables

Table 6.3 summarises the Pearson's bivariate partial correlations between measures for the whole sample, controlling for age. One-tailed significance testing was used. Note that the zero-order correlations are not reported here as they only varied by from the partial correlations by up to ± .03.

Overall a similar pattern of correlations occurred at outcome (41-49 months) as was noted at initial assessment (24-30 months). This indicates stability of relationship between these variables between ages two and four years. The results support the hypothesis that aspects of working memory and language are associated at ages 41-49 months. All aspects of working memory tested here correlated with the PLS-4 scores to an alpha level of .05; except for PLS-4 EC with Inhibit. Receptive vocabulary showed a lower degree of correlation with working memory measures, only meeting this significance level for the TENR, KWM II and Working Memory. The TENR and KWM II correlated to a moderate degree, *r* (70) = .37, lower than the large association seen at initial assessment, *r* (75) = .50. EFs and processing

Table 6.3

Bivariate correlations between the variables at 41-49 months (N = 78)

Variables	1.	2.	3.	4.	5.	6.	7.	8.	**9.**	**10.**	**11.**	**12.**	**13.**	**14.**	15.
1. PLS-4 EC raw score	_														
2. PLS-4 AC raw score	.91***	_													
3. PLS-4 total raw score	.98***	.97***	_												
4. ROWPVT raw score	.75***	.74***	.76***	_											
5. VRO raw score	.63***	.57***	.61***	.45***	_										
6. DEAP – PCC	.54***	.41***	.49***	.40***	.38***	_									
7. KWM II total score	.62***	.65***	.65***	.56***	.23*	.25*	_								
8. TENR total score	.68***	.59***	.65***	.54***	.50***	.67***	.37***	_							
9. Inhibit (log)	-.17†	-.21*	-.19*	-.04	.04	-.09	-.18†	-.03	_						
10. Shift (log)	-.42***	-.29**	-.38***	-.17†	-.08	-.31**	-.29**	-.28**	.39***	_					
11. Emotional Cont. (log)	-.32**	-.32**	-.34**	-.17†	-.01	-.22*	-.30**	-.18†	.66***	.60***	_				
12. Working Memory (log)	-.31**	-.31**	-.32**	-.22*	-.15	-.09	-.37***	-.22*	.74***	.42***	.63***	_			
13. Plan / Org. (log)	-.30**	-.25*	-.29**	-.18†	-.10	-.12	-.31**	-.16†	.70***	.52***	.67***	.85***	_		
14. GEC (log)	-.36***	-.34**	-.36***	-.19†	-.08	-.20*	-.35**	-.21*	.87***	.64***	.85***	.89***	.90***	_	
15. LWL II (ms)	-.27**	-.27**	-.29**	-.12	-.18†	-.29**	-.20*	-.20*	.25*	.10	.34**	.19†	.16†	.27*	_

Note. Items in bold (9-14) are from the BRIEF-P assessment. PLS-4 AC / EC = The Preschool Language Scale, (Fourth Edition) (PLS-4) Auditory Comprehension / Expressive Communication; ROWPVT-4 = The Receptive One Word Picture Vocabulary Test (Fourth Edition); VRO = The Mullen Early Scales of Learning: Visual Reception Organisation; DEAP - PCC = Diagnostic Evaluation of Articulation and Phonology Percentage of Consonants Correct; KWM II = The Key Word Working Memory II task; TENR = Test of Non-Word Repetition; GEC = Global Executive Composite; LWL = Looking While Listening II task mean latency.
†p<.10. * p<.05. **p<.01. ***p<.001.

speed correlated to a small to moderate level. The strength of relationship between the TENR and the PLS-4 EC decreased from the initial assessment to outcome, but remained a large effect, *r* (75) = .81 to *r* (72) = .68. The LWL II mean latency no longer had a large correlation with VRO scores; *r* (67) = -.60 at initial assessment compared with *r* (73) = -.18 at outcome. It continued to show a small association with language, and small to moderate association with EF measures. Another development is that a closer relationship between the PLS-4 subtests was noted in this study (the correlation between PLS-4 EC and PLS-4 AC increased from *r* (76) = .87 to *r* (74) = .91); possibly because the disparity between receptive and expressive language had resolved for most late talkers over the 18 month period.

6.3.2 Effect of demographic variables on expressive language at 41-49 months

Expressive vocabulary (CDI scores) was used as the dependent variable in Study 1, however for Study 2, the CDI was no longer an age-appropriate measure. I was unable to add an age-appropriate measure of expressive vocabulary to the test protocol due to time constraints. Therefore the PLS-EC raw scores were used as the measure of expressive language skills in this study. As for Study 1, the degree of influence of the demographic variables needs to be established so those with a significant effect on expressive language can be included in the statistical models. This allows a clearer picture of the unique impact each processing skill has on language development.

A series of univariate analyses of covariance (ANCOVA) were run to estimate the effect of demographic variables on children's expressive language (PLS-4 EC) scores, controlling for age, at the outcome assessment. Age at the time of outcome assessment was not correlated with PLS-4 EC scores, *r* (76) =.03, *p* = .39, one-tailed. However, it was controlled for in all the following analyses, in case there were a disproportionate number of older or young children in any of the groupings evaluated below.

Boys had a lower score for PLS-4 EC scores (*M* = 53.69, *SE* = .97) than girls (*M* = 57.88, *SE* = 1.45) when age was controlled, $F(1,74) = 5.68$, $p = .02$, partial eta squared = .07. The mean differences in PLS-4 EC scores was 4.18 (*SE* = 1.75), 95% CI [.69, 7.73].

Children with a positive family history of speech language or learning difficulties in first or second degree relatives had a lower average PLS-4 EC scores (*M* = 53.60, *SE* = 1.24) than children who did not (*M* = 56.11, *SE* = 1.10) when controlling for age; however the difference was not significant, $F(1,74) = 2.28$, $p = .14$, partial eta squared =.03.

The effect of birth order on PLS-4 EC scores was evaluated. First-born (*M* = 55.51, *SE* = 1.08), second-born (*M* = 55.26, *SE* = 1.44) and third-born children (*M* = 54.61 (*SE* = 3.20) scored nearly the same, however fourth-born children (*M* = 46.04, *SE* = 4.15) scored lower. However the difference in means was not significant, $F(3,72) = 1.65$, $p = .19$, partial eta squared = .06.

Many families had another child born between the initial assessment and outcome visits. This change in environment may have affected the children's language development. However there was no clear pattern of mean scores relating to the number of children in the family at outcome. When controlling for age, participants who were the only child in the family had a mean PLS-4 EC score of 54.93 (*SE* = 1.94); those who were one of two children (*M* = 55.93, *SE* = 1.07); three children (*M* = 53.79, *SE* = 1.70); four children (*M* = 36.09, *SE* = 7.00) and five children (*M* = 51.09, *SE* = 4.96). The model was not significant, $F(4,70) = 2.29$, $p = .07$, partial eta squared = .12. Pairwise comparisons revealed that the differences between groups arose because of significant differences between four children in the family and one, two or three. None of the other pairwise comparisons were significant. One child who was diagnosed with a language disorder at outcome was one of four children; his unusually low score may have skewed these results. Numbers of families with four or five children in the family were very low, providing a further reason to not to over interpret these results.

Only children (reported at ages 24-30 months) continued to score higher on expressive language at outcome (M = 56.16, SE = 1.46) than children-with-siblings (M = 54.44, SE = 1.01). However the difference in means was not significant, $F(1, 74) = .92$, p = .34, partial eta squared = .01. This shows that sibling status at initial assessment no longer had a significant effect on PLS-4 EC scores by the outcome assessment. Even children who still remained singletons at the outcome assessment scored essentially the same on the PLS-4 EC at outcome (M = 54.93, SE = 2.03), as those-with-siblings (M = 55.02, SE = 0.91), $F(1, 74) = .001$, p = .97, partial eta squared = .00.

Children were grouped by their parent's highest qualification level. Six groups were formed based on the New Zealand Qualifications Authority (NZQA) framework (as shown in Table 3.1). The children's PLS-4 EC scores showed a significant difference as a function of parent education level, $F(5,69) = 5.26$, $p < .001$, partial eta squared = .28.

Partial correlations were calculated for the interval data while controlling for age at outcome assessment. A one-tailed test was conducted for the birth-weight and weeks-premature variables. Birth-weight was not significantly correlated with PLS-4 EC scores ($r(73) = .08$, $p = .24$). Prematurity also was not significantly correlated ($r(75) = -.05$, p = .33). Hours in day-care at outcome assessment was not significantly correlated with outcome PLS-4 EC scores ($r(72) = -.15$, $p = .20$, two-tailed).

Summary of the effect of demographic variables on expressive language

A different pattern of significant associations was found with the outcome expressive language scores and demographic variables compared with the results of Study 1. Firstly, age did not correlate significantly with the scores at ages 41-49 months. Age does impact on language scores, but the age range tested was narrow and its effect could not be seen in this group. As in the initial assessment, boys still scored lower than girls. This result may be an artefact of the loading of the sample in Study 1 with late talkers (68% of whom were boys). A positive family history of language, speech or learning difficulties was not significantly associated with lower expressive language in this sample. A trend was shown for children from larger families to have poorer language scores at outcome; however the difference was not significant and the small number of children with family sizes of four and five children means this trend should not be over interpreted. The difference in expressive language between only-children and those-with-siblings (both at initial assessment and those who were still the only-child at outcome) had become statistically insignificant by age 41-49 months. Finally, there was a stronger association between highest parent qualification and PLS-4 EC scores at outcome than at initial assessment. This may reflect the cumulative effect of an enriched linguistic environment over time for those with parents with higher education. Birth weight, prematurity and hours in day care at outcome assessments were not significantly associated with PLS-4 EC scores at outcome assessment.

6.3.3 Initial multivariate regression model for concurrent expressive language at 41-49 months

Having identified the significant bivariate associations with expressive language outcomes both for concurrent behavioural variables and demographic factors, the next step was to determine which of these contributed unique variance to a multivariate model. Following the same statistical methods as for Study 1, the best model for predicting concurrent expressive language was established using a mixture of hierarchical and backwards linear regression.

Age was entered into the multiple regression model first and left in as Step 1 in subsequent iterations, even although it was not a significant predictor at the $p<.05$ level. The predictors which significantly associated with PLS–4 EC on a bivariate level were entered on the second step: sex, parent qualifications, ROWPVT-4, PCC, KWM II, LWL II mean latency, TENR, Shift (log), Emotional Control (log), Working Memory (log) and Plan/ Organise (log). These were removed one by one, in order of least significance, until only significant predictors remained in the model. The final model is summarised in Table 6.4, $F(6,64) = 38.99$, $p < .001$.

Table 6.4

Multiple regression analysis summary for concurrent variables predicting expressive language (PLS-4 EC raw scores) at 41-49 months (N = 74)

Model	β	SE β	Standardised β
(Constant)	55.95	11.26	
Age	-.39	.21	-.11†
Parent Quals	1.09	.33	.21**
KWM II	.15	.05	.23**
TENR	.09	.03	.27**
Shift (log)	-15.7	4.86	-.20**
ROWPVT-4	.23	.05	.40***

Note. Parent Quals = Highest Parent Qualification; KWM II = Key Word Working Memory Task II ; TENR = Test of Non-Word Repetition; ROWPVT-4 = Receptive One Word Picture Vocabulary Test (Fourth Edition). †p<.10. * p<.05. **p<.01. ***p<.001.

This model accounted for 79% of the variance in PLS-4 EC raw scores. The amount of unique variance accounted for holding all others constant was age 1%; parent qualifications 4%; KWM II 3%; TENR 5%; Shift (log) 4% and ROWPVT 8%.

See Appendix F for the diagnostics procedure which was followed for this linear regression model.

Summary

These results show partial support of the hypothesis that aspects of working memory predict expressive language at ages 41-49 months. There are some differences in the associations amongst variables compared with the patterns observed at 24-30 months. Receptive vocabulary was the most powerful predictor (8% unique variance) at 41-49 months, whereas it had not been a significant predictor of expressive vocabulary at 24-30 months. The TENR predicted much more unique variance in expressive vocabulary scores (22%) than the KWM I scores (3%) at ages 24-30 months (see Section 4.3.2.2), but they were more similarly associated with expressive language scores (5% and 3% respectively) by ages 41-49 months. I investigated whether this change could be due to the change in dependent variables between Studies 1 and 2 (CDI cf. PLS-4 EC scores). The best regression model predicting PLS-4 EC scores at 24-30 months using concurrent measures was nearly identical to the one predicting CDI scores (Table 4.4). The TENR predicted 23.9% unique variance in this model and the KWM I task predicting 6.4%. (The main difference between these models was that only-child status was not a significant predictor of PLS-4 EC scores). Therefore these results at 41-49 months seem to reflect a different balance of associations between PSTM / VWM and expressive language over time. This could reflect the developmental shift from emerging lexicons to developing more advanced language forms over this time period.

Shift accounted for a similar amount of unique variance as VWM and PSTM (3.5%) at ages 41-49 months, again affirming its unique association with language development in this age range. Parent qualifications were a significant predictor of unique variance at 41-49 months, whereas they had not been at 24-30 months. Only-child status however did not predict unique variance at 41-49 months, whereas it had at 24-30 months. These changes likely reflect the shift in environmental influences on children's language development over time as has been previously reported (Reilly et al., 2010). Once again, processing speed and the other EFs did not contribute unique variance in multivariate regression models.

6.4 Research question 6

Do these aspects of working memory (PSTM, processing speed, VWM, VSWM and EF) improve on previously established multivariate linear regression models for predicting expressive language at 41-49 months?

Research hypothesis

Working memory measures will improve previously established multivariate linear regression models predicting expressive vocabulary at ages 41-49 months.

6.4.1 Comparison multivariate regression model for concurrent expressive language at 41-49 months

A second multiple linear regression analysis predicting concurrent expressive language was then completed using a forwards hierarchical method. The purpose of this was to investigate the degree to which the working memory predictors added to existing predictive models of concurrent expressive language. As in Study 1 (see Section 4.5.2), in this analysis the variables were forced into the model in order of how well their association with concurrent expressive language had been previously established in the literature in this age range. Unlike in Study 1, there are no studies similar enough to this one to use as a direct comparison, however age, parent qualifications, receptive vocabulary and non-word repetition have all been previously established as predictors of concurrent expressive language at ages 41-49 months (Reilly et al., 2010; Thal et al., 2005). Therefore age was entered into the model first into the model, followed by the ROWPVT-4, parent qualifications, then the TENR. Next the KWM II and finally Shift measures were entered to determine if these new predictors contributed unique variance over and above the previously established variables. Table 6.5 shows the results of this series of analyses.

The final model (Step 6) was significant, $F(6,64) = 38.99$, $p < .001$. The predictors accounted for 79% of the variance. Age was not a significant predictor in any of the models. All the other predictors were significant at the $p < .05$ level. When entered hierarchically in this order, age contributed less than 0.1% unique variance, ROWPVT-4 56%, parent qualifications 6%, TENR 8%, KWM 5% and Shift (log) 4%.

Table 6.5 Hierarchical multiple regression analysis summary for concurrent variables predicting expressive language (PLS-4 EC raw scores) at 41-49 months (N = 74)

Step and predictor variable	β	SE β	Standardised β	R^2	ΔR^2
Block 1				.001	.001
Constant	49.98	18.98			
Age	.11	.42	.03		
Block 2				.56	.56
Constant	35.39	12.75			
Age	-.14	.28	-.04		
ROWPVT-4	.43	.05	.75***		
Block 3				.62	.06
Constant	34.87	12.00			
Age	-.18	.27	-.05		
ROWPVT-4	.40	.05	.69***		
Parent Quals	1.29	.41	.24**		
Block 4				.70	.08
Constant	39.77	10.80			
Age	-.40	.24	-.11		
ROWPVT-4	.30	.05	.52***		
Parent Quals	.92	.38	.17*		
TENR	.12	.03	.36***		
Block 5				.75	.05
Constant	35.59	9.98			
Age	-.38	.22	-.11†		
ROWPVT-4	.22	.05	.38***		
Parent Quals	1.09	.35	.21**		
TENR	.11	.03	.32***		
KWM II	.19	.05	.28**		
Block 6				.79	.04
Constant	55.95	11.26			
Age	-.39	.21	-.11†		
ROWPVT-4	.23	.05	.40***		
Parent Quals	1.09	.33	.21**		
TENR	.09	.03	.27**		
KWM II	.15	.05	.23**		
Shift (log)	-15.7	4086	-.20**		

Note. ROWPVT-4 = Receptive One Word Picture Vocabulary Test (Fourth Edition); Parent Quals = Highest Parent Qualification; TENR = Test of Early Non-Word Repetition; KWM II = Key Word Working Memory II task. †$p < .10$. * $p < .05$. **$p < .01$. ***$p < .001$.

Summary

This additional analysis demonstrates that the KWM II and Shift measures contributed unique variance (5% and 4% respectively) to expressive language scores at ages 41-49 months over and above previously established predictors (receptive language 56%; parent qualifications 6% and non-word repetition 8%). This finding supports the hypothesis that working memory variables will add predictive capability to previously established models of expressive language on a group level.

6.5 Research question 7

What differences in group means can be seen between the RLT and TD groups in working memory skills (PSTM, processing speed, VWM, VSWM and EF) and language at ages 41-49 months?

Research hypothesis

Performance on measures of working memory will be lower in RLTs than in TD children, particularly in PSTM.

6.5.1 Comparison of group means at 41-49 months between RLT and TD groups

Children were regrouped according to their language outcomes. Children with a standard score on either PLS-4 AC or EC subtests of below 85 were classed as “LI” (*n* = 5). There were no exclusionary criteria therefore this group included children with both “non-specific” and “specific” LIs.

Interviews with their parent(s) confirmed that these children were demonstrating substantial functional difficulties with expressive and or receptive language in the clinic setting and at home relative to what was typical for their age. Three sets of parents were concerned about their child’s language and or communication skills and had already initiated accessing speech language therapy services for their children (P17, P30 and P53). Two of these children had been recently diagnosed with autism (P30 and P53) and the other child (P17) had low standard scores in visual cognition (VRO score = 74) and both subscales of the PLS-4 (AC = 81 and EC = 78) at outcome. The remaining two sets of parents whose children met these criteria did not indicate concerns about language or communication while talking with the researcher or in the Parent Questionnaire II.

The first child (P69) whose parent did not indicate concern had a standard score of 70 on the PLS-4 EC subscale and was just beginning to combine words at age 43 months. This parent did not have any formal school qualifications and commented that she had difficulties with learning. This child had recently been referred to local speech language therapy services by another health professional who had been in contact with the family.

The final child (P79) whose parent did not indicate concern had a low PLS-4 AC standard score at initial assessment (81) and at outcome (83), whereas his expressive language scores were in the typical range at both initial (103) and outcome assessment (94). His mother did not think he had difficulty understanding language, but did comment that "he wasn't good at listening" and that she had to repeat instructions to him frequently. There was some caution about including this child in the LI group however, as his parents reported they spoke a second language at home 20-30% of the time. This may have meant that the standard scores on the PLS-4 underestimated his true language ability. However I was also concerned about his VWM capacity. He could only follow two key word instructions at his outcome assessment e.g. "the cat wants the drink", and his scores for Working Memory on the BRIEF-P at both initial and outcome assessments were approximately 1.5 standard deviations above the mean for the sample (27 and 26 points respectively). This showed a high level of problem behaviours related to poor working memory capacity in everyday situations as perceived by his mother. I was concerned that if his difficulties with VWM and receptive language were persistent, that this could place him at a significant disadvantage for academic achievement over time (Gathercole et al., 2005). Therefore he was included in the LI group. The remaining 73 children were either classed as being in the RLT or TD groups depending on their group status in Study 1.

Due to the small numbers in the LI group (*n* = 5), comparing them statistically to the RLT and TD groups was not feasible. However the RLT and TD groups were compared with each other in order to identify any factors which could aid prediction of resolution of early language delay. The mean age of the RLT group at follow up was 44.40 (*SD* = 2.09) months and the TD group was aged 44.94 (*SD* = 2.00) months on average. This difference was not significant, t (71) = 1.02, p = .31. The time between assessments was not significantly different between groups, t (71) = -0.09, p = .93; the TD group had an average of 18.23 months (*SD* = 1.09) between assessments and the RLTs had 18.25 (*SD* = 0.85).

Table 6.6 summarises the mean, standard deviation and range of scores of the behavioural assessments for the TD, RLT and LI groups. Tests of statistical significance are reported for the TD and RLT group comparisons. *T*-tests were used to compare the assessments with a normal distribution. The EF measures from the BRIEF-P were compared using Mann-Whitney *u* tests, as their distributions departed significantly from normal (see Table 6.2 for skew / kurtosis values). Cohen’s *d* values were used as a measure of effect size. Any *d* values over 0.8 were interpreted as a large effect, over 0.5 a moderate effect and over 0.2 a small effect size (Cohen, 1988).

Table 6.6

Comparison of RLT, TD and LI groups results for the behavioural assessments

	Typically developing N = 53			Resolved late talkers N = 20			Language impairment N = 5			Statistical tests for differences of means (RLT / TD)		
Measures	N	Mean (SD)	Range	N	Mean (SD)	Range	N	Mean (SD)	Range	t-test	d	Mann-Whitney
PLS-4 EC (SS)	52	124.79 (11.51)	94-148	20	110.65 (9.87)	94-135	5	74.40 (16.15)	50-94	4.85 ***	1.32	
PLS-4 AC (SS)	53	120.45 (11.48)	90-142	20	110.05 (11.11)	93-133	5	75.80 (14.79)	50-87	3.48***	.92	
PLS-4 total (SS)	52	125.02 (11.90)	97-150	20	111.55 (10.85)	94-135	5	73.40 (13.83)	50-87	4.40***	1.18	
ROWPVT (SS)	53	115.92 (9.26)	95-141	20	106.50 (9.75)	95-134	4	90.00 (3.367)	85-92	3.82***	1.10	
VRO (SS)	53	116.85 (13.42)	85-145	18	109.61 (13.37)	82-143	4	81.25 (8.85)	74-92	1.98*	.54	
TENR – raw score	51	106.53 (13.58)	65-127	20	82.35 (22.44)	18-113	4	60.25 (23.47)	38-82	5.55***	1.30	
KWM II – total raw score	51	51.12 (8.61)	22-60	19	44.47 (11.47)	21-58	4	31.25 (10.01)	19-43	2.62**	.67	
DEAP - PCC	53	83.85 (11.74)	42-100	20	67.50 (13.28)	35-89	4	61.75 (21.17)	38-82	5.12***	1.30	
BRIEF-P:												
Inhibit (raw score)	52	19.69 (4.56)	16-33	19	19.89 (4.23)	16-29	4	19.75 (4.50)	16-26		-.05	476.00
Shift (raw score)	52	11.79 (2.35)	10-20	19	13.95 (4.08)	10-27	4	14.00 (2.94)	10-17		-.65	303.00**

Measures	Typically developing N = 53			Resolved late talkers N = 20			Language impairment N = 5			Statistical tests for differences of means (RLT / TD)		
	N	Mean (SD)	Range	N	Mean (SD)	Range	N	Mean (SD)	Range	t-test	d	Mann-Whitney
Emotional Control (raw score)	52	12.63 (2.66)	10-22	19	14.68 (4.27)	10-25	4	14.00 (2.71)	12-18		-.58	353.50*
Working Memory (raw score)	52	19.50 (4.01)	17-34	19	21.16 (5.98)	17-36	4	21.50 (3.11)	19-26		-.33	443.50*
Plan / Organise (raw score)	52	11.87 (2.50)	10-18	19	13.79 (4.32)	10-24	4	11.50 (1.29)	10-13		-.54	377.50†
GEC – total raw Score	52	75.25 (13.19)	63-113	19	83.47 (20.31)	63-124	4	80.75 (9.78)	72-94		-.48	358.50*
LWL II task (ms)	53	0.50 (0.06)	0.37-0.64	20	0.51 (0.07)	0.43-0.68	4	0.53 (0.13)	0.40-0.67	-.23	-.15	

Note. PLS-4 AC / EC = The Preschool Language Scale, (Fourth Edition) (PLS-4) Auditory Comprehension / Expressive Communication; ROWPVT-4 = The Receptive One Word Picture Vocabulary Test (Fourth Edition); VRO = The Mullen Early Scales of Learning: Visual Reception Organisation; TENR = Test of Non-Word Repetition; KWM II = The Key Word Working Memory II task; DEAP - PCC = Diagnostic Evaluation of Articulation and Phonology Percentage of Consonants Correct; GEC = Global Executive Composite; LWL = Looking While Listening II task mean latency.
†p<.10. * p<.05. **p<.01. ***p<.001.

Informal comparisons of the three groups' mean language scores support the view that children with LI form the lower end of the language ability spectrum, with RLTs performing better than this group, then finally children who talked on time performing at the top end of the range. Due to the small number of children with LI at outcome (*n* = 5), this analysis focuses on group differences between the RLTs and TD children. After a period of 18 months, the two groups were significantly different on all measures except Inhibit, Plan / Organise and LWL II mean latency. The RLT group scored lower on expressive and receptive language and receptive vocabulary than the TD group with large effect sizes. There was nearly a whole standard deviation of difference between the two groups on expressive language. Large group differences were also seen in performance on the TENR. Speech sound production (DEAP - PCC) lagged behind in the RLT group with a large effect size. There were significant differences in KWM scores as in Study 1, although the RLTs fell behind with a moderate effect size in Study 2, rather than the previously large effect in the late talker group. In terms of the EF measures, group differences were significant on Shift with a moderate effect size, as was seen as in the late talkers at age two years. The RLTs also were also reported to have more problem behaviours with Emotional Control than TD children, also with a moderate effect size and Working Memory showed a small effect in the same direction. The GEC total score showed a significant increase in problem behaviours for the RLT group with a nearly moderate effect size. Finally there was no detectable difference in processing speed on the LWL II task between the two groups, unlike the significant finding of a moderate group effect between late talkers and TD children at age two years.

6.6 Research question 8

Do any significant differences in working memory skills (PSTM, processing speed, VWM, VSWM and EF) remain when language differences between the RLT and TD groups are controlled?

Research hypothesis

These effects are expected to be seen even when controlling for group differences in language skills.

6.6.1 Comparison of group means in working memory while controlling for language at 41-49 months

To check that the differences in working memory skills in the RLT group were not just a result of their lower language levels, a series of ANCOVAs was run to explore whether group differences held even, when the most relevant language skill(s) to the working memory construct under investigation were controlled. This measures investigated were: TENR, KWM II, Shift, Emotional Control and Working Memory.

As for Study 1 (Section 4.6.1), TENR scores were predicted using group status (RLT / TD) and concurrent receptive vocabulary scores and expressive phonology as covariates. The model was significant, $F(1,67) = 5.88$, $p = .02$, partial eta squared = .08. This is interpreted as a medium effect size. The mean estimated difference between groups was 11.11 points (SE = 4.58), which equates to eleven extra phonemes correctly repeated by the TD group on average over the assessment, over and above the effects of receptive vocabulary and expressive phonological development. These results indicate that PSTM abilities were fairly stable across the 18 month period between assessments for this group, as the late talker group also scored lower than the TD children on the TENR when controlling for receptive vocabulary and phonology scores in Study 1. This provides additional support for the hypothesis that poor PSTM might have been a factor in the RLT group's early expressive language delays as it continues to be particularly low in this group, even once their expressive language skills resolved.

An ANCOVA comparing group means for KWM II task between RLT and TD groups while controlling for PLS-4 AC raw scores was not significant. The groups scored similarly on the KWM II task at outcome once PLS-4 AC scores were taken into account, $F(1,67) = .98$, $p = .33$, partial eta squared = .01. This indicates that there is no support for the hypothesis that group differences in VWM measured at 41-49 months were any greater than expected given their differences in receptive language abilities.

The assumption of homogeneity of variance was met for the ANCOVA comparing KWM II scores, but not for the ANCOVAs comparing the TENR group means. However the assumption of normality of residuals held for these analyses, and a linear model was an appropriate fit for these data. ANCOVA has been shown to be a robust analysis when only one of these assumptions is violated (Rheinheimer & Penfield, 2001). This was not the case for the ANCOVA models comparing Shift, Emotional Control and Working Memory scores across the two groups. The high number of zero scores on each of these EF subscales resulted in heavily skewed distributions and linear modelling was not appropriate for these data. Transforming the data did not improve model fit. Non-parametric options to compare the effect of receptive language ability on groups differences in these scores were explored (e.g. removing participants with zero scores), but no suitable solutions were available which retained all the data points. Therefore these three variables (Shift, Emotional Control and Working Memory) were not examined in relation to Research Question 4.

Summary

The hypothesis that RLTs would have lower working memory scores (particularly in PSTM), even when language ability was controlled, was partially supported. The RLT group had significantly lower scores on the TENR, KWM II, Shift, Emotional Control and Working Memory scores. However the RLT group also had lower scores in the language measures with large effect sizes. Even when group differences in receptive vocabulary and expressive phonology had been taken into account, group differences in TENR scores at age 41-49 months remained significant. This provides further support for the view that capacity limitations in PSTM may have constrained early expressive vocabulary growth in the RLT group, as the group weakness in PSTM was still evident after their expressive language delays had resolved. However there was no evidence to support the view that group differences in VWM (measured by the KWM II task) were any greater than expected, given the RLTs' weaker receptive language skills. Due to difficulties with the distribution of scores in the BRIEF-P measures, Shift, Emotional Control and Working Memory were not able to be examined in relation to this question.

6.7 Chapter summary

Overall, the working memory measures have shown a similar pattern of association with language and late talking (now resolved) as they did in Study 1. PSTM, VWM and Shift were once again the only working memory variables to contribute unique variance in the multivariate linear regression model predicting concurrent expressive language scores. The repetition of this finding suggests general stability in the relationships between these constructs and expressive language across ages two to four years.

In this cohort, PSTM (measured by the TENR) was significantly lower in RLTs compared with TD children at ages 41-49 months, even when receptive vocabulary and expressive phonology were controlled. These results support the view poor PSTM is a key feature of late talkers, even once resolved, and may be a factor in their early delays.

The KWM task proved to be a useful measure of VWM from ages two to four years. This task contributed 5% unique variance over and above previously established predictors of expressive language for this age group, namely age, receptive vocabulary, parent qualifications, sex and non-word repetition. However while VWM was lower in the RLT group compared with the TD group, this effect was no longer seen once the group differences in receptive language abilities were controlled. This suggests that the lower VWM skills displayed by the RLT group were associated with their lower receptive language skills, and does not provide evidence to support the hypothesis that additional deficits in VWM may contribute to early expressive language delays.

The relationship between EF and language ability showed a similar pattern results at age 41-49 months as seen at 24-30 months. As in Study 1, Shift was found to account for unique variance (4%) in expressive language scores, over and above previously established predictors of expressive language and VWM. Despite a pattern of significant correlations with language measures on a bivariate level, none of the other BRIEF-P measures (Inhibit, Working Memory, Emotional Control or Plan / Organise) were predictors of unique variance in the multivariate regression model for concurrent expressive language. Group differences in the three EF measures which were significantly lower in the RLT group compared with the TD group (Shift, Emotional Control and Working Memory) may have been attributable to group differences in receptive language, however this hypothesis was not tested, as linear models were inappropriate for predicting group differences in BRIEF-P subscale scores and alternative analyses removed too many cases to be informative.

Finally, as in Study 1, processing speed was found to associate with expressive language at age 41-49 months, but not to predict unique variance once working memory measures were in the models. There was no significant difference in processing speed between the RLT and TD groups, despite the lower language scores of the RLT group. This finding will be considered further in the discussion chapter.

The final study will report on the clinical outcomes of this cohort at 41-49 months. The question of whether aspects of working memory measured at ages 24-30 months improve prediction of total language outcomes at ages 41-49 months over previously established predictive models will be explored on both group and individual levels.

Chapter Seven: Study 3 Results

7.1 Research question 9

7.1.1 Clinical outcomes at 41-49 months of age

7.2 Research question 10

7.2.1 Bivariate associations between the measures at 24-30 months and PLS-4 total language scores at 41-49 months

7.2.2 Initial multivariate regression model for total language scores at 41-49 months using variable measured at 24-30 months

7.3 Research question 11

7.3.1 Comparison multivariate regression model for total language outcomes at 41-49 months using variables measured at 24-30 months

7.4 Research question 12

7.4.1 Individual prediction

7.5 Chapter summary

7.1 Research question 9

What clinical outcomes were evident at ages 41-49 months?

Research hypothesis

50%-75% of late talkers are expected to resolve and score within the normal range on total language scores on the PLS-4 by ages 41-49 months. Due to the liberal criteria for late talker status, no former TD children are expected to have a LI.

7.1.1 Clinical outcomes at 41-49 months of age

Table 7.1 summarises the outcomes of the cohort at ages 41-49 months of age. Outcomes are presented for language disorders, low visual cognition, low phonology, diagnosis of autism, clinician and parent concerns.

The hypothesis was partially supported. As previously stated in Study 2, five children were classed as LI at outcome. Four of these children had been late talkers and unexpectedly, one was initially classed as TD (P79). The prediction that the liberal cut-off used for classifying "late talkers" in this study would preclude TD children being in the LI group at outcome was not upheld. Eight three percent of the late talkers (20/24) had resolved their language delay by age 41-49 months. This rate of resolution was higher than predicted. The use of the PLS-4 as the outcome measure may have inflated this figure. Multiphasic measures may underestimate those having specific difficulties in morphosyntax which is known to lag behind in children with resolving early language delays (Paul & Alforde, 1993).

Overall, the main linguistic weakness seen in the late talker group at ages 41-49 months was poor phonology. 70% of the late talkers scored one standard deviation below the mean or more for phonology (measured by PCC on the DEAP) (see Table 7.1). However, the severity of these errors should not be overestimated. Fifty seven percent of the late talkers had no delayed error patterns and 74% had no unusual error patterns. This figure therefore

Table 7.1 Summary of clinical outcomes at 41-49 months for the whole group, LT and TD groups (groups defined at ages 24-30 months)

Outcome measure	Total sample (N = 78) % (N)	Typically developing (N = 54) % (N)	Late talkers (N = 24) % (N)
Contact with SLT services by 41-49 months:			
No assessment or intervention	86 (67)	98 (53)	58 (14)
Assessment only and or monitoring	8 (6)	2 (1)	21 (5)
Has received intervention	6 (5)	-	21 (5)
Low VRO score (<85)	6 (5)	2 (1)	17 (4)
Low PLS-4 score (<85 on one subtest)	6 (5)	2 (1)	17 (4)
Low PCC (< standard score 7)	33 (26)	19 (10)	70 (16)
Delayed error patterns:			
None	80 (62)	91 (49)	57 (13)
1	10 (8)	6 (3)	22 (5)
2+	9 (7)	4 (2)	22 (5)
Unusual error patterns:			
None	87 (68)	94 (51)	74 (17)
1	8 (6)	6 (3)	13 (3)
2+	4 (3)	-	13 (3)
Diagnosis of autism	3 (2)	-	9 (2)
Clinician concern:*			
No concerns	71 (55)	83 (45)	42 (10)
Monitor	13 (10)	13 (7)	13 (3)
Refer for SLT assessment / intervention	17 (13)	4 (2)	46 (11)
Area of clinician concern:			
No concerns	71 (55)	83 (45)	42 (10)
Phonology only	19 (15)	15 (8)	29 (7)
Language only	1 (1)	2 (1)	-
Phonology and Language	3 (2)	-	8 (2)
Complex needs	6 (5)	-	21 (5)
Parent concerns:			
Hearing	6 (5)	6 (3)	9 (2)
Language	18 (14)	6 (3)	48 (11)
Communication	12 (9)	6 (3)	26 (6)

Note. VRO = Mullen Visual Reception Organisation; PLS-4 = Preschool Language Scales, Fourth Edition; PCC = Percentage of Consonants Correct; SLT = Speech language therapy. * Ratings of clinical concern at were made by an experienced therapist immediately after the outcome assessments at ages 41-49 months.

reflects a pattern of *poor* phonology, rather than delayed or disordered, in the majority of late talkers.

Decisions regarding clinical concern were made immediately after the outcome assessments were completed. Weaknesses in any aspect of speech production, language, voice or fluency were taken into account in deciding whether or not to recommend monitoring or a referral for speech language therapy (SLT) assessment and or intervention. These clinical decisions were made on a holistic basis, taking into account functional communication, parent concern, environmental factors and any other developmental issues. The child's language status at two years of age and the steps parents had taken to address any concerns were also considered when making recommendations. Finally recommendations for a referral to SLT services were made without reference to the availability or affordability of these services; instead the potential benefit to the child was the main factor in this decision.

Clinical concern at 41-49 months was elevated in the late talker group (14/24) compared with the TD group (9/54). To determine whether this difference was significant, a chi-square test was performed. As two cells had expected values less than five, the table was collapsed into a binary clinical outcome: 'no concerns' or 'monitor / refer' (Field, 2009). The difference between groups was significant, χ^2 (1, N = 78) = 13.87, p < .001. This represents the finding that based on the odds ratio, children identified as late talkers at age two years were seven times more likely to be monitored or referred for speech language therapy services 18 months later than initially TD children, 95% CI [2.37, 20.65]. Late talking at age two years was a risk factor for clinical concern at age 41-49 months in this cohort.

The clinical outcomes at age 41-49 months of each late talker are described in brief below. There were no clinical concerns regarding 10/24 former late talkers. For the remaining 14, concern over poor phonological skills alone accounted for seven of these rating of clinical concern. Five former late talkers were identified with complex needs[5] (difficulties in more than one aspect of development). Two of these had been diagnosed with autism since the initial assessment (P30 and P53). One former late talker (P6) was now stuttering and presented with disordered phonology, and the final two scored less than one standard deviation below the mean on at least one of the PLS-4 subscales and also for visual cognition (Mullen VRO) (P17 and P69). The final two former late talkers (P3 and P82) attracted clinical concern as they had a combination of delayed phonology and weak functional communication skills (although they scored in the typical range for language on the PLS-4).

Forty two percent of the late talker group had been seen by SLT services for assessment, monitoring or intervention between the initial and outcome assessments, compared with only 2% of the TD group. It is difficult to interpret these rates, as after the initial assessment was completed for each child, I discussed each child's results with their parents and talked through the options regarding local SLT services. Some parents of late talkers requested advice about how to support their child's language development at home and this was briefly given. These discussions would have affected parents' decisions on whether or not to seek additional assessment for their children. In other words, if they had not participated in this study, the rates of access to SLT services may have been different.

Only three children (P30, P53 and P17) were referred to SLT services by myself (with parent permission) after the initial assessments of this study. These children presented with complex needs at the initial assessment. At 41-49 months all three met the criteria for LI. The remainder of the parents of late talkers were advised to monitor their children's progress and to contact myself for advice or refer directly to local SLT services if they were still concerned

[5] Note that all the children were retained in the sample for Studies 1-3 (even those with general delays or autism) as it is common for children to change diagnostic categories over time (see Section 1.1.1).

after three to six months. Therefore these children accessed SLT services at their parents' discretion between the initial and outcome assessments of this study.

By 41-49 months, only five former late talkers had received intervention for speech and language delays. Even within this small group, the amount and type of intervention varied widely. One of these children (P27) had received eight weeks of twice weekly intervention for an expressive language delay, and had resolved his delay by the outcome assessment. At the other end of the extreme, P53 (who had autism) had a multidisciplinary team working with him (an SLT, psychologist, Early Intervention Teacher and an Educational Support Worker (paraprofessional)). He was receiving six hours per week of support worker time in his early childhood centre to provide one-on-one support to reach his language and social goals. He also had fortnightly visits from the SLT and Early Intervention Teacher to work with him, monitor his progress and provide on the job training for his support worker. This level of support had been in place for approximately 12 months. This child still met the criteria for a LI at the outcome assessment. These two cases illustrate that the role of intervention in the resolution of late talkers identified at age two years is difficult to determine due to the wide variety of access to intervention services, and large differences in the underlying severity of individual cases of delay. However as a simple summary, 20/24 of the late talkers resolved their language delays, and of these only three had received SLT intervention before the outcome assessment. Intervention therefore did not play a large role in the high resolution rates seen in this cohort of late talkers.

7.2 Research question 10

What are the bivariate and multivariate associations between total language scores on the PLS-4 at 41-49 month olds and the following aspects of the working memory system: PSTM, processing speed, VWM, VSWM and EF) measured at 24-30 months?

Research hypothesis

Significant associations will be seen between all aspects of early working memory skills and outcome total language scores on a bivariate level. Working memory skills will predict unique variance in the multivariate model, the strongest predictor being PSTM.

7.2.1 Bivariate associations between the measures at 24-30 months and PLS-4 total language scores at 41-49 months

PLS-4 total language standard scores were used as the outcome measure, as they summarised the overall language ability of the child better than either the EC or AC subtests alone. The overall strength of the child's linguistic system is more relevant to clinical outcomes than their expressive language outcomes alone as receptive language delays are more resistant to intervention than expressive delays and therefore are generally considered more of a concern (Law, Garrett & Nye, 2004).

Table 7.2 shows the Pearson's bivariate partial correlations (controlling for age at initial assessment) between the outcome PLS-4 total standard score and the assessment data from the initial assessments for the whole sample. One-tailed significance testing was used. The complete table is not shown here as correlations amongst the initial assessments have already been presented in Table 4.3 (Study 1).

These correlations support the hypothesis that aspects of working memory measured at 24-30 months will be associated on a bivariate level with language outcomes 18 months later. All but the A Not B task, Inhibition and Plan / Organise were significantly associated with total language standard scores. Large effect sizes were seen for associations between language outcomes and initial KWM I and TENR scores. A moderate association was evident between early processing speed and later language total scores. Small effects were present between early Shift, Emotional Control, GEC and later language total scores.

Table 7.2

Bivariate partial correlations between PLS-4 total standard scores at 41-49 months and the behavioural assessments at 24-30 months

Initial assessment variables (24-30 months)	Outcome PLS-4 total language standard scores (41-49 months)
CDI (words produced)	.66***
TENR (raw score)	.58***
KWM I (raw score)	.77***
A not B total score	.09
ROWPVT-4 (raw score)	.75***
PLS-4 AC raw score	.78***
PLS-4 EC (raw score)	.75***
VRO (raw score)	.61***
TPT-PCC	.64***
LWL I mean latency (log) (ms)	-.44***
BRIEF-P:	
Inhibit (log)	-.11
Shift (log	-.29**
Emotional Control (log)	-.30**
Working Memory (log)	-.27*
Plan / Organise (log)	-.14
GEC (log)	-.25*

Note. CDI=Communicative Development Inventory; TENR = Test of Non-Word Repetition; KWM = The Key Word Working Memory task; ROWPVT-4 = The Receptive One Word Picture Vocabulary Test (Fourth Edition); PLS-4 AC / EC = The Preschool Language Scale, (Fourth Edition) (PLS-4) Auditory Comprehension / Expressive Communication; VRO = The Mullen Early Scales of Learning: Visual Reception Organisation; TPT- PCC = Toddler Phonology Test – percentage consonants correct; LWL I = Looking While Listening I task; GEC = Global Executive Composite.
†p<.10. * p<.05. **p<.01. ***p<.001.

7.2.2 Initial multivariate regression model for total language scores at 41-49 months using variables measured at 24-30 months

The demographic and behavioural variables which were significantly associated with total language standard scores on the PLS-4 were entered into a linear regression model and removed one by one beginning with the least significant one, until only significant predictors remained in the model. These predictors were sex, parental education (degree status), TENR, KWM I, PCC, VRO (raw scores), LWL I mean latency (log), Shift (log), Emotional Control (log), Working Memory (log). I decided to use PLS-4 AC and EC raw scores as predictors rather than the vocabulary measures, as they had higher bivariate associations with the outcome variable. Age at initial assessment and time between assessments (measured as months between the CDI date and first outcome assessment session date) were included in the model, even although they only reached significance levels when other variables were included. The final model, comprising only significant predictors, is summarised in Table 7.3.

Table 7.3

Multiple regression analysis summary for variables measured at 24-30 months predicting PLS-4 total language standard scores at 41-49 months (N = 76)

Model	β	SE	Standardised β
(Constant)	186.08	32.18	
Time between assessments (months)	-2.59	1.06	-.15*
Age at initial assessment (months)	-2.90	.54	-.33***
Parent education	4.44	2.24	.12*
KWM I	1.24	.21	.52***
Shift (log)	16.99	7.60	.19*
Emotional Control (log)	-20.86	6.44	-.27*
PLS-4 AC raw score	.97	.24	.37***

Note. * p<.05. **p<.01. ***p<.001.

This model was significant, $F(7,67) = 32.82$, $p < .001$, and accounted for 77% of the variance in PLS-4 total language standard scores. The amount of unique variance accounted for holding all others constant was time between assessments 2%; age at initial assessment 10%; parent education 1%; KWM I 12%; Shift (log) 2%; Emotional Control (log) 4% and PLS-4 AC standard scores 5%.

See Appendix F for the diagnostics procedure which was followed for this linear regression model.

Summary

This result is broadly in line with the research hypothesis that working memory skills at age two years will be associated with total language outcomes 18 months later. The variable contributing the largest amount of unique variance to total language scores at outcome was the KWM I task (12%). This new finding supports the view that VWM may play a role in early language acquisition during this developmental period. Children with more initial problems in the area of Emotional Control (as measured by the BRIEF-P) had slightly worse language outcomes. This could be seen to support the hypothesis that weaknesses in the working memory system (in this case, the CE) contribute towards poorer language outcomes. Alternatively this result could be interpreted as evidence of increased early frustration on the part of children with poorer language outcomes. Poorer Shifting at 24-30 months (which also involves the CE) was associated with better total language outcomes at 41-49 months in this regression model. This was an unexpected finding given that poorer Shifting was associated with worse concurrent expressive language at both time points tested. This apparent change in the direction of influence between Shift and language may indicate that this is a spurious result. This will be considered further in the discussion chapter. Finally, none of the other measures of the CE (Inhibit, Plan / Organise, Working Memory or the A not B task) at 24-30 months were related to total language outcomes.

In confirmation of previous studies (Chiat & Roy, 2008; Thal et al., 1991), early receptive language was a predictor of later total language outcomes, here contributing 5% unique variance. This finding likely reflects the common pattern for late talkers to have relatively better receptive language at intake and then to resolve their expressive language delays over time. Parental education was also a significant predictor of a small amount of unique variance (1%), as has been previously reported in a similar study (Reilly et al., 2010).

I hypothesised that PSTM would be the strongest predictor of later language outcomes, due to its well established association with LI in the literature. However early TENR scores (measuring PSTM) predicted later total language scores on a bivariate level but did not predict unique variance in the multivariate models. To explore this, a regression analysis was run, predicting total language standard scores at outcome using the early TENR and early receptive language scores. The model was significant, $F(2, 74) = 40.74$, $p < .001$, and accounted for 52% of the variance in total language scores at outcome. Early receptive language was a significant predictor ($p < .001$), but the TENR was not ($p = .40$). Therefore the lack of unique association of early PSTM with later language outcomes is because of the shared variance with early receptive language. Similar results were seen when early expressive language was used as a predictor instead of receptive language. These findings will be considered further in the discussion.

Early processing speed, despite a moderate correlation with later language outcomes, did not predict unique variance in the multivariate regression model. This appears to contradict the findings of Marchman and Fernald (2008). Their study had a different age of outcome (eight years cf. 41-49 months), yet the age of initial assessment and measures of expressive vocabulary and processing speed were very similar. They did not include measures of working memory in their initial assessment, but did include them as part of their outcome measures. To determine whether the current results contradict or extend their findings, their regression analysis was repeated using this cohort's data. Expressive vocabulary (CDI words produced) and processing speed (LWL I task) measured at 24-30 months were used to predict PLS-4 expressive language standard scores at 41-49 months using the enter method in a linear regression analysis. The model was significant, $F(2, 67) = 24.08$, $p < .001$, and accounted for 42% of the variance in expressive language scores. Both predictors were significant at an alpha level of .05. The KWM I task was then entered into the model as a third predictor. In this model, $F(3,66) = 27.70$, $p < .001$, early processing speed was no longer significant ($p = .75$); while the CDI and KWM I were significant predictors ($p < .01$). This model accounted for 56% of the variance in expressive language scores at outcome. Therefore the current study extends Marchman and Fernald's (2008) findings. Early processing speed only contributed unique variance to language outcomes 18 months later, when VWM was not included in the model. The caveat here is that the age that language outcomes were measured at was quite different (41-49 months cf. eight years). Replicating Marchman and Fernald's (2008) outcome assessment battery at age eight years with the current cohort would allow for a more robust comparison between these two studies.

7.3 Research question 11

Do these aspects of working memory (PSTM, processing speed, VWM, VSWM and EF) improve on previously established multivariate linear regression models for predicting group language outcomes at 41-49 months of age?

Research hypothesis

Aspects of working memory measured at 24-30 months are expected to improve established predictive models of group total language outcomes at 41-49 months.

7.3.1 Comparison multivariate regression model for total language outcomes at 41-49 months using variables measured at 24-30 months

The next step was to investigate the extent to which the working memory predictors improved on previously established predictive models of language. A multiple linear regression analysis with the same dependent and independent variables was completed using a forward hierarchical method. In this analysis, the predictors were entered into the model in order of the degree to which their influence on language outcomes had been previously established in the literature. Receptive language has been well established as a moderate predictor of language outcomes measured at three to four years in late talkers (Ellis Weismer, 2007; La Paro et al., 2004; Lyytinen et al., 2001; Roulstone et al., 2003). Parent education has been found to be a significant predictor in two large scale longitudinal studies of late talkers, albeit with a small effect size (Dale et al., 2003; Reilly et al., 2010). There have been no previous studies of VWM, Emotional Control or Shift as predictors in this age range. Therefore receptive language was entered into the model as the first step, followed by parent education. These were followed by KWM, Emotional Control and Shift scores (the new variables were entered in order of beta size from the enter regression model). Table 7.4 shows the results of this series of analyses.

Table 7.4

Hierarchical multiple regression analysis summary for variables measured at 24-30 months predicting PLS-4 total language standard scores at 41-49 months (N = 76)

Step and predictor variable	β	SE β	Standardised β	R^2	ΔR^2
Step 1				.59	.59***
Constant	44.06	7.40			
PLS-4 AC SS	.68***	.07	.77		
Step 2				.61	.02†
Constant	39.71	7.62			
PLS-4 AC SS	.64***	.07	.72		
Parent education	5.35*	2.81	.15		
Step 3				.67	.08***
Constant	47.01	7.09			
PLS-4 AC SS	.39***	.09	.43		
Parent education	5.99*	2.54	.17		
KWM I	.94***	.23	.40		
Step 4				.69	.01
Constant	70.88	17.02			
PLS-4 AC SS	.35	.09	.40		
Parent education	5.43	2.54	.15		
KWM I	.98	.22	.42		
Emotional Control (log)	-8.18	5.32	-.11		

Step and predictor variable	β	SE β	Standardised β	R^2	ΔR^2
Step 5				.71	.02*
Constant	48.35	19.70			
PLS-4 AC SS	.38***	.09	.43		
Parent education	4.98*	2.49	.14		
KWM I	1.05***	.22	.44		
Emotional Control (log)	-18.26*	7.03	-.24		
Shift (log)	18.06*	8.49	.20		

Note. PLS-4 AC SS = Preschool Language Scales (Fourth Edition) Auditory Comprehension standard scores; KWM I = Key Word Working Memory Task I. * $p < .05$. ** $p < .01$. *** $p < .001$.

This series of regression analyses shows that when entered in this order, initial receptive language scores predicted 59% of unique variance in outcome PLS-4 total language scores; parent education predicted 2%, KWM predicted 8%; Emotional Control predicted 1% and finally Shift predicted 2%.

Summary

This hierarchical regression analysis demonstrated that the new KWM I task measured at 24-30 months adds unique predictive variance (8%) to previously established predictive models of language outcomes at 41-49 months of age (early receptive language and parent education). Emotional Control and Shifting also accounted for small amounts of additional unique variance (1-2%) in total language scores when both were in included the model, but not when entered on their own. This may indicate the association between these two EF measures and later language is less robust than the other predictors' associations. Overall however this analysis confirms the hypothesis that measures of working memory at 24-30 months would improve predictive models for group total language outcomes measured at ages 41-49 months.

7.4 Research question 12

Do working memory measures at 24-30 months (PSTM, processing speed, VWM, VSWM and EF) improve prediction of total language outcomes at 41-49 months for individual late talkers?

Research hypothesis

The addition of early working memory variables into predictive models is expected to improve prediction for individual late talkers.

7.4.1 Individual prediction

There were only five children who were classified as LI at the outcome assessment. This number was too small for logistic regression analyses. However an informal analysis of patterns for these five children demonstrates the challenge of individual prediction of LI status well. Of the four late talking children who met criteria for LI at outcome, only one clear pattern can be seen from the initial assessment data: An inability to engage in any clinician-led standardised assessment at ages 24-30 months could be a risk factor for autism. Two children in this cohort were not able to engage in any such assessments and both were diagnosed with autism by age 41-49 months.

The results of the multivariate regression analyses pointed to a unique association between early VWM and later total language scores at a group level; however this task did not help predict outcomes on an individual level. All the children with LI at outcome had scores below five on the KWM I task initially, as did three other late talkers who resolved. The same pattern could be seen for exceptionally low CDI, VRO, TENR, A not B scores, slowed processing speeds and BRIEF-P scores.

The second robust predictor of unique variance in later language identified at a group level was early receptive language. All five children in the LI group had standard scores of < 85 on the PLS-4 AC at 24-30 months. However, so did three other late talkers who resolved their language delays. Therefore early low receptive language did not always result in LI 18 months later.

If children with early receptive language delays (PLS-4 AC standard score <85 at 24-30 months) had been selected for monitoring for poor language outcomes, rather than late talkers, nine children would have been identified in Study 1: (P3, 17, 30, 49, 53, 60, 69, 77, 79). All five children later diagnosed with LI would have been in this group, including the child who was initially classed as TD (P79). Of the four children with early receptive language delays who were not in the LI group at outcome, two had accessed SLT services for speech and or language concerns (P3 and P60), and the remaining two (P49 and P77) had resolved their difficulties without accessing any clinical services. This represents a resolution rate of 44.4% for language and 22.2% for having no clinical concerns in the early receptive language delay group. Children with early receptive language delays were 12 times more likely to attract clinical concern (‘monitor’ or ‘refer’) 18 months later than those whose comprehension scored within the normal range at 24-30 months, $\chi^2(1, N = 78) = 11.41$, $p = .001$, 95% CI for the odds ratio [2.19, 61.46].

Summary

To summarise, the hypothesis that working memory skills measured at 24-30 months would improve individual prediction of total language outcomes at 41-49 months was not supported. This seems likely to be due to the heterogeneity of children’s developmental patterns. None of the assessments here, either alone or in combination, consistently distinguished between clinical groups at outcome for every individual. However children with early receptive language delays were at increased risk of LI than late talkers, and 7/9 of them had accessed speech language therapy services by age 41-49 months. Due to the small number of children with early receptive language delays in this cohort, these results may not generalise to the wider population.

7.5 Chapter summary

This cohort demonstrated a high degree of resolution of late talking compared with similar studies. Only five children had a LI at outcome and of these four had been late talkers initially. Regression models show that in this cohort, age, time between assessments, receptive language, VWM, Emotional Control and Shift at 24-30 months predicted total language outcomes 18 months later to a strong degree (77% of variance in PLS-4 total language standard scores).

As hypothesised, VWM, Emotional Control and Shift improved prediction of language outcomes on a group level above previously established models. VWM was the most powerful addition to previous models, contributing an additional 8% unique variance. This finding supports the view that VWM is associated with language acquisition across this developmental period. Shift and Emotional Control were only significant predictors of later language outcomes (adding 2 and 1% unique variance respectively) when both were present in the model. These two predictors are therefore viewed with caution. Further difficulties arise in interpreting the results for Shift. The nature of the relationship between Shift and language scores reversed in the longitudinal study (Study 3), compared with the two concurrent studies (Studies 1 and 2). In Studies 1 and 2, more difficulty with Shifting was associated with lower language, whereas in Study 3 more difficulty with Shifting at 24-30 months predicted better language outcomes 18 months later. This may indicate a spurious result. These findings will be considered in the discussion. Also unexpectedly, PSTM did not predict unique variance in total language outcomes over time in the multivariate models; PSTM was more strongly associated with the initial stages of vocabulary acquisition as demonstrated in Studies 1 and 2.

Despite the good fit of the regression model for group language outcomes, none of the significant variables in this model improved prediction of individual outcomes. The two best predictors, VWM and receptive language, did not improve prediction of outcomes on an individual level. However in this cohort at least, children with early receptive language delays would have been better group to monitor for LI than late talkers. Late talkers appeared to be more at risk for phonological concerns at 41-49 months than LI; late talkers having a higher incidence of both delayed and disordered speech processes compared with TD children. This finding raises questions about the links between early poor PSTM, delayed expressive vocabulary and phonological development which could be explored in further studies.

Chapter Eight: General discussion

8.1 Introduction

8.2 Phonological Short Term Memory (PSTM)

8.3 Processing speed

8.4 Summary of findings for the Central Executive (CE)

8.4.1 Verbal working memory (VWM)

8.4.2 Visual spatial working memory (VSWM)

8.4.3 Executive functions (EF)

8.4.3.1 Behaviour Rating Inventory of Executive Functions – Preschool Version (BRIEF-P)

8.4.3.2 Shift

8.4.3.3 Emotional Control

8.4.3.4 Inhibition

8.4.3.5 Working Memory

8.4.3.6 Plan / Organise

8.4.3.7 Summary of Executive Functions (EF) and language

8.5 Models of individual prediction

8.6 Limitations of the current research

8.7 Strengths of the current research

8.8 Implications for clinical practise

8.9 Future research

8.10 Conclusion

8.1 Introduction

The practical question which inspired this research was: “Is there anything wrong with my late talking toddler?” There were two overarching hypotheses proposed to be explored in this thesis. Firstly, that deficits in working memory were implicated in late talking. Secondly that adding working memory measures to existing models would improve predictions of language outcomes at four years of age. Three studies were presented which were designed to explore these hypotheses. In Study 1, I examined the concurrent relationships between measures of the working memory system, visual cognition and language in TD and late talking in children aged 24-30 months. Group comparisons between the late talkers and TD groups were made to determine whether aspects of working memory were associated with early language delays. Study 2 repeated this methodology, after an 18 month period, to examine the stability of relationships between these variables over time. Children who were resolved late talkers (RLTs) were compared with the TD group to see whether aspects of working memory could account for their pattern of early delay followed by resolution of language skills. Finally in Study 3, I reported the clinical outcomes of the cohort. I examined whether aspects of working memory measured at 24-30 months improved predictive models for language outcomes at 41-49 months of age in this cohort. Finally I considered whether early working memory measures allowed for accurate prediction of LI status at 41-49 months on an individual level. A discussion of the results is presented below. The overall pattern of the data were considered for evidence of a main direction of influence between working memory and lanaguge in the age range two to four years, while acknowledging no firm conclusions about casuality could be drawn from this study design. This chapter concludes with a comment on the strengths and weaknesses of the study, implications for clinical practise and suggested directions for future research.

8.2 Phonological Short Term Memory (PSTM)

This study has extended previous findings of associations between expressive vocabulary and non-word repetition reported by other studies in the age range 20-36 months (Chiat & Roy, 2007; Hoff et al., 2008; Stokes & Klee, 2009b). As hypothesised, Studies 1 and 2 showed the TENR scores (measuring PSTM) were a powerful predictor of concurrent expressive vocabulary skills (correlations of $r(77) = .80$ with expressive vocabulary at 24-30 months and $r(73) = .68$ with expressive language at 41-49 months). The TENR also predicted significant unique variance in expressive scores at both time points in multiple regression analyses (8% in expressive vocabulary scores at ages 24-30 months and 5% in expressive language scores at age 41-49 months). These results support prior research which proposes a role for PSTM in early vocabulary acquisition (Gathercole & Baddeley, 1990; MacRoy-Higgins et al., 2013; Rescorla, 2009; Stokes, 2013; Stokes et al., 2012; Vernes et al., 2008; Whitehouse et al., 2011).

Furthermore, the results suggest PSTM may play a stronger role in the early stages of expressive vocabulary acquisition rather than later language development. Both late talkers (at 24-30 months) and RLTs (at 41-49 months) had significantly lower TENR scores than TD children, even when language and phonology differences between the groups were controlled. This shows an ongoing weakness in PSTM in the RLT group at ages 41-49 months even once early language delays have resolved. As mentioned in the literature review, there is evidence for bidirectional links between working memory and language and the main direction of influence between them may change over time (Gupta & Tisdale, 2009). Metsala's (1999) account of the relationship between non-word repeitition and vocabulary predicts a corresponding improvement in non-word repetition alongside improvements in language over time, driven by improvements in phonological awareness. However these data are more consistent with the view that poor PSTM constrained early expressive vocabulary acquisition and may have even been a causal factor in some cases of late talking.

According to the capacity limits theory, poor PSTM would slow early vocabulary acquisition as more repetitions of a new word would be needed before a detailed enough phonological representation was stored to allow that word to enter the child's lexicon. In order to produce a word expressively, more detailed representations are needed than to recognise it (for receptive vocabulary). Therefore this theory predicts that poor PSTM would constrain expressive vocabulary development more than receptive (Gathercole, 2006; Munroe et al., 2012). This theory is reflected in the data of the current study, in that late talkers and RLTs both had relatively better receptive vocabulary than expressive at both time points, though still scoring significantly behind the TD groups.

In Study 3, it was hypothesised that children with better early PSTM skills would have better language outcomes. However while early TENR scores were significantly associated with total language outcomes on a bivariate level, they were not a significant predictor in the multivariate model once early receptive (or expressive) language was also entered, assumedly due to shared variance between these measures. This again confirms the strong relationship between early TENR scores and early language development.

Despite the fact that the TENR did not predict unique variance in later language outcomes in this cohort, all the children with LI at outcome had poor initial TENR scores. Bishop (2006) suggested that the more processing deficits a child has, the more likely it is that a clinical disorder will arise. In the current cohort it seems the RLT group was able to compensate for their poor PSTM skills over time, whereas the five children who were diagnosed with a LI were not. Therefore it is likely that the children with LI had additional difficulties, which caused a longer lasting constraint on their development of language. This view is consistent with the findings of Bishop, Adams and Norbury (2006). They reported on the language outcomes of 173 six-year-old twin pairs who were identified with an early risk of LI at age four years. Their DeFries-Faulker analysis showed that while poor non-word repetition and poor verb tense development both had genetic origins, they seemed to have different genetic origins. This study indicates that poor early non-word repetition would not necessarily be a useful predictor of later difficulties with general language skills, as was the case in the current research.

However, these findings contrast with those of Chiat and Roy (2008), who found that early real and non-word repetition was the strongest predictor of morphosyntax outcomes at ages four-to-five years. However there are multiple differences between their study and the current one. Firstly their sample was older at intake (30-42 months) and more severe than the current late talker group, the children all having been referred for clinical services. The outcome measures were also different, with the current research using a multiphasic measure based on total language standard scores (which combined receptive and expressive standard scores) as the outcome variable rather than a measure of morphosyntax. It is also possible that floor effects in the late talker group in the current research at ages 24-30 months may have reduced the long term predictive capability of the TENR.

To further support this finding that poor PSTM skills are a characteristic of late talkers, three other studies have reported significant weaknesses in non-word repetition skills in RLTs compared with TD children (Bishop et al. (1996); D'Odorico et al. (2007); Thal et al. (2005)). Petruccelli et al. (2012), however, failed to find evidence to support the view that RLTs have poor non-word repetition compared with TD peers. While they found a trend towards non-word repetition deficits in their late talkers, the group differences in their study were not statistically significant. A comparison of these four studies showed differences in age of identification of language delay, criteria for delay, controls for phonological delays, criteria and ages for identifying LI and sample sizes. The three studies which found statistically significant group differences all had smaller late talker sample sizes (*N* = 13-20), than Petruccelli et al.'s study (*N* = 45). Larger studies are less likely to encounter type I errors, which indicates Petruccelli et al.'s study cannot be disregarded easily. The most likely cause of the discrepant findings is that in Petruccelli et al.'s sample, the RLTs did not have significantly lower language or non-verbal cognition than the TD children. This is not typical of other late talker cohorts. Another possible reason is that Petruccelli et al.'s sample was representative of a range of SES backgrounds. It is possible that in a wider cross-section of children, more environmental influences on language acquisition are introduced. The role of environmental effects on a developing linguistic system has not been investigated in relation to late talkers or the role of PSTM in language acquisition. However it is an important variable to be considered given Metsala's (1999) prediction that increased vocabulary will improve phonological awareness which will allow improved non-word repetition. Therefore, the conflicting results of these four studies may reflect the heterogeneity of causes for late talking. This is an avenue for future research.

Studies 1 and 2 documented higher correlations between non-word repetition and concurrent expressive vocabulary (r = .80 and .68 respectively) than has been previously documented in the published literature (the range reported in other studies covered here is r = .28 - .60). Three reasons for these differences will be discussed: methodological differences, the developmental level of the children and sample composition. Firstly, similar to the current research, Stokes and Klee (2009b) examined the relationship between expressive vocabulary and PSTM in children aged 24-30 months using the TENR (an earlier version). In Study 1 of the current research, the TENR was more highly associated with CDI scores, $r(77) = .80$, than in Stokes and Klee's (2009b) study, $r(178) = .60$. In Stokes and Klee's study, the children who did not complete the TENR (23% of their sample) were excluded; a decision which likely excluded many late talkers. Not only did the current research include children with partial scores, but those who were not able to imitate words yet were given a zero score; thus capturing the full range of ability on the TENR. In addition, I used a later version of the TENR which included five syllable words. Therefore greater variance in both ends of the ability spectrum in non-word repetition was captured in the current research. Higher correlations are expected to result when this is done.

The second reason for the higher correlations between non-word repetition and vocabulary in the current research could be the developmental stage of the children in the current study (24-49 months). Acquisition of everyday vocabulary is a strong feature of linguistic development in this age range. The majority of developmental research in non-word repetition however, studied children aged four years and older. While large gains in vocabulary are still occurring over this time period, a substantial lexicon has already been established which may act as a foundation for future gains, reducing reliance on PSTM in new word learning (Gupta & Tisdale, 2009). In the current research it seems that poor PSTM only constrained vocabulary acquisition intitally, with the late talkers accelerating their development assumedly once a bank of vocabulary had been acquired. Studies of children in the age range four to six years have reported correlations ranging from r = .3 - .5 (Bowey, 2001; Gathercole et al., 1992). Gathercole et al. (1992) reported a lower correlation by age eight years (r = .28). These studies used different measures than the current research, which means they are not directly comparable. However according to the capacity limit model, higher correlations between non-word repetition and vocabulary could be most reasonably expected in the age range where expressive vocabulary is emergent, as is reflected by the data cited here. Finally, the sample in the current research had a higher proportion of later talkers (approximately 30%) than would be naturally occurring in the general population (approximately 15%) and assumedly than were present in Stokes and Klee’s (2009) sample.

The loaded sample may also have contributed towards a higher association between PSTM and expressive vocabulary.

Finally the finding of poorer phonological skills in the late talker and RLT groups in Studies 1 and 2 compared with the TD groups raises two important questions. As non-word repetition relies on intact auditory perception, phonological processing and speech-motor output processes as well as phonological storage (Gathercole, 2006), late talkers' poorer TENR scores may simply be a by-product of poor speech-motor planning and output processes rather than a genuine difficulty with memory. Several arguments can be made to support the assumption that poor TENR scores reflect poor PSTM. Firstly the TENR was designed for children in this age range and only 11% of its phonemes are later developing sounds (/r/, /l/, /s/ and /ʃ/). There are no consonant clusters. Under the current scoring procedures if a child who was gliding (a common developmental process in this age range) produced a single instance of the target liquid in the TENR or concurrent phonology test (TPT or DEAP), they were scored incorrect for gliding that phoneme on TENR items. This scoring procedure was based on psycholinguistic theory which predicts that non-word repetition does not access a stored form, but instead goes straight from input to output. Therefore if a child is able to articulate a phoneme, they should be able to repeat it accurately in a non-word. More lenient scoring of currently developing phonemes may have slightly attenuated the group differences in TENR scores. However, this is unlikely to account for the group differences in TENR scores as these remained significant even when phonology and language were added as covariates on the ANCOVA group comparisons at both age ranges tested. Some caution in interpreting the results of Study 1 are warranted given the floor effects on both the TENR and TPT for the late talkers. However, this was not an issue in the RLT group at ages 41-49 months, therefore indicating that the group differences in the TENR scores appear to be robust differences over and above differences in the group's phonological development. In further support of this, two similar studies have reported that more conservative scoring of their non–word repetition tests made little or no difference to their results (Gray, 2003; Roy & Chiat, 2004).

Having established this, the second question can be considered: What is the nature of the link between poor PSTM and phonological delays? There is little research linking the development of phonology with PSTM at this stage, although some authors are beginning to link together theories of speech processing models with those of WM, such as the work by Jacquemot and Scott (2006). It is possible that poor early PSTM constrains the development of phonology, although once again causality could be bidirectional. This is another avenue of future research.

8.3 Processing speed

The hypothesis that processing speed would be associated with language at both age ranges tested was partially supported. Processing speed (as measured by the LWL I task) showed significant associations with language on a bivariate level at 24-30 months. However processing speed did not predict language skills in either of the multiple regression analyses (predicting concurrent expressive vocabulary at 24-30 months or predicting total language scores at 41-49 months from processing speed measured at 24-30 months). These findings extend the work of Marchman and Fernald (2008) who reported that processing speed (as measured by a LWL task) and expressive vocabulary measured at 25 months of age predicted 41% variance in expressive language scores (on the Expressive Language Index of the Clinical Evaluation of Language Fundamentals – Fourth Edition; Semel, Wiig and Secord, (2003)) in the same cohort at eight years of age. In the current research, the variance that early processing speed (measured at ages 24-30 months) contributed to later expressive language scores (at ages 41-49 months) was no longer significant once early KWM I scores (measuring VWM) were added to the model. Therefore working memory seems to mediate the effect of processing speed on language. Processing speed is thought to be a subcomponent of the working memory system which impacts on the efficiency of memory processing, and thus length of span (Bayliss et al., 2005). In contrast, Poll et al. (2013) in their study of TD primary school aged children found that while processing speed contributed unique variance to working memory capacity, it also accounted for additional unique variance in language performance when the task was taxing. Also Leonard et al. (2007) found that working memory and speed were separable constructs in models predicting concurrent language in 14 year olds with LI and their TD peers. There are large differences between these two studies and the current research (e.g. the ages of the children, severity of language difficulties, range and type of measures, statistical comparisons etc.), which means they are not directly comparable. However they illustrate the point that the relationship between speed and language is likely to be more complex than the current research indicates, for example it could vary by the age and severity of the children, and the measures used. At this stage, the lack of unique associations between speed and language measures in the current research suggests that early working memory measures are more useful in predicting language outcomes in toddlers than processing speed.

There were expected to be significant group differences in processing speeds between the late talker and TD groups at 24-30 months and the RLT and TD groups at 41-49 months of age. These hypotheses were supported at 24-30 months, but not at 41-49 months of age. This was despite the significant large group differences in language and working memory skills between the RLT and TD groups at outcome. This may mean that RLTs do not have slower spoken word recognition than TD children. However there may be other reasons for this finding. The psychometric properties of the LWL I and II tasks must be considered given the discrepancy in findings. The two tasks were the essentially comparable (if not exactly the same); the exceptions being the technology used to present the tasks and collate the data, and the ages of the children. The eye tracker used for the LWL II task allowed for more data points to be gathered per child, and at a faster sampling rate; however there was no reason to lose confidence in the methods used to gather data in the LWL I task, which are well established (Fernald et al., 2008). Instead the likely cause of this dissonance is the developmental changes in the children over the 18 month period. The instruction “look at the <target>” may not have elicited the simple reaction to shift eye gaze at ages 41-49 months, that it did at 24-30 months, but instead triggered a series of thoughts about the target and distractor pictures, followed by an eye gaze shift. This would mean the task no longer functioned as a reaction time measure. In support of this view, the mean whole group latencies were slower in Study 2 compared with Study 1, whereas there is evidence to expect processing speed for language tasks to increase with age throughout the preschool years (Fernald, Pinto, Swingley, Weinberg, & McRoberts, 1998). In further support of this view, the distribution of the LWL II mean latencies was normal, rather than the positive skew typical of reaction time data. With hindsight, the LWL II task does not seem to be suitable for measuring reaction time in children aged 41-49 months and therefore the results from it should be viewed with caution. Conclusions about processing speed are therefore restricted to the results of the LWL I task measured at ages 24-30 months.

8.4 Summary of findings for the Central Executive (CE)

8.4.1 Verbal Working Memory (VWM)

On a methodological level, the KWM task is assumed to have been successful at measuring VWM in two to four year old children. Concurrent validity was demonstrated by the pattern of correlations with the other assessments measuring similar constructs in the study, such as the PLS-4 AC ($r(75) = .73$ at 24-30 months and $r(70) = .65$ at 41-49 months) and the TENR ($r(75) = .50$ at 24-30 months and $r(70) = .37$ at 41-49 months). Results from the comparison regression analysis in Study 3 showed that VWM at 24-30 months contributed unique variance to total language scores at 41-49 months, even once early receptive language was entered into the model first. The KWM task therefore measures an ability to follow verbal instructions of increasing memory load which is closely aligned with receptive language and PSTM, but distinct from them, purportedly, VWM.

In Studies 1 and 2, I hypothesised that there would be significant associations between VWM and expressive vocabulary / language skills at both time points measured. In Study 3, I predicted that better VWM at 24-30 months would predict better total language outcomes at 41- 49 months. These hypotheses were supported by the data. At ages 24-30 months, the KWM I task predicted a small amount of unique variance in concurrent expressive vocabulary scores in the best regression model (1%). A similar pattern emerged at age 41-49 months, when the KWM II task predicted 3% unique variance in expressive language scores. The KWM I task measure at 24-30 months also predicted 12% unique variance in total language scores at ages 41-49 months over and above the variance predicted by other variables, which included early receptive language. This pattern of results supports the view that VWM plays a unique role in language development for the whole group, particularly over time. Between the ages of two and four years, a large amount of language acquisition is occurring. This includes learning basic syntax and morphology, everyday gains in vocabulary and the development of early narrative, discourse and pragmatic skills. These skills thereafter form the foundation of every utterance and linguistic interaction. The results of this study indicate that a larger capacity for VWM gives an advantage in language acquisition in this developmental period.

In Studies 1 and 2, I hypothesised that there would be differences in VWM skills in the late talkers and RLT groups compared with the TD groups at both time points, even when controlling for group differences in receptive language, but these hypotheses were not supported by the data. While VWM was shown to be poorer in late talkers at 24-30 months than TD children, this group difference was no longer significant once concurrent receptive language was controlled. This finding was replicated in the RLT and TD groups at ages 41-49 months. VWM is thought to play a role in language development, therefore controlling for receptive language would also remove some of the variance in VWM itself. There is circularity in proposing the direction of causality between VWM and language, indicating that the results of the ANCOVAs where language was controlled need to be interpreted with caution. Therefore these null findings do not disprove the view that late talkers and TD children or RLTs and TD children may differ in VWM. However given that such group differences were found in TENR scores when controlling for receptive vocabulary and phonology (at both age ranges tested), but were not seen for VWM, it seems reasonable to assume that poor PSTM is more characteristic of children with early expressive language delay than poor VWM.

These are new findings in this age range. Most published work on the role of VWM in first language acquisition has used concurrent or cross-sectional samples in primary school aged children. The few preschool studies including VWM and vocabulary tests in their assessment batteries were focused on EF rather than language development, such as Hughes (1998) and Willoughby et al. (2010). The findings of the current study are in line with research linking VWM skills and sentence comprehension for novel or complex syntax in older TD children and children with SLI (Magimairaj & Montgomery, 2012a, 2012b; Montgomery & Evans, 2009) and also with the studies in other populations which have linked VWM with later language outcomes in the upper primary school years (Kronenberger et al., 2013; Pierpont et al., 2011; Pisoni et al., 2011). This is an exciting trend in the literature and warrants further corroborating evidence.

8.4.2 Visual Spatial Working Memory (VSWM)

The rationale for including a VSWM measure in this research was as a non-verbal measure of the CE. This assumed the domain generality of the CE functions, as per Baddeley's model (2007). I hypothesised that there would be associations between VSWM and language at both age ranges tested. Group differences were also expected between the late talker and TD groups at 24-30 and RLT and TD groups at 41-49 months of age. Contrary to expectations, at ages 24-30 months, the A not B task did not correlate with any other task in Study 1, except Shift and then only to a small degree. The A not B task at 24-30 months was also not a significant predictor of total language outcomes at 41-49 months. This was despite evidence of CE involvement in language development through the KWM I task results. Despite revisions to extend the A not B task for the outcome assessments at ages 41-49 months, strong ceiling effects were observed in Study 2, and it was dropped from the analysis. Therefore this discussion of the A not B task is based on the results of Studies 1 and 3 only. There are two possible interpretations of the null findings in Studies 1 and 3. Firstly, it is possible the A not B task was not valid and reliable enough to measure VSWM accurately in this cohort. As mentioned in the Chapter 3, this task appeared to have adequate psychometric properties, except that test-retest reliability had only been reported to be moderate (Epsy et al., 2008). According to this interpretation, with hindsight, perhaps I should have developed a novel preschool measure of VSWM with better psychometric properties to use in this research, instead of using the A not B task. This novel task could have been designed to be comparable to the structure of the KWM task, and thus may have better supported the aims of this study.

A second interpretation of these null findings is also possible; in this case assuming that the A not B task did measure VSWM adequately. VSWM was included in this study under the assumption of the domain generality of the CE (Baddeley, 2007). Under this assumption, EFs are categorised by function (e.g. shifting / inhibition / updating), rather than domain (e.g. verbal / spatial). Since this study began, this model of the CE has been questioned. Nee et al. (2013) published a meta-analysis of fMRI studies which investigated patterns of neural activation in EF tasks. The analysis indicated that two frontal regions were activated in EF tasks: The caudal superior frontal sulcus was particularly activated by spatial information ("where" information), and the mid-lateral prefrontal cortex was activated with non-spatial verbal and object information processing ("what" information). The authors therefore suggested that a dual model of EF (a "where" / "what" framework) fitted neuroimaging data better than a model based on function. According to this new framework, behavioural tasks which are based on the same type of content should correlate more highly with each other, than those which differ in content, regardless of which particular EF is under scrutiny. This leads us to reconsider the results from the A not B task in the current research. According this alternative model, the A not B task could have been successfully measuring EF in the spatial domain. Nee et al.'s framework predicts a lack of correlations between the A not B task and language, as they cross content areas ("where" versus "what" respectively). In further support of this view, several studies have reported that the A not B task is an appropriate task for measuring EF in the spatial domain for the ages range two to five years (Epsy et al., 1999; Epsy et al., 2001; Griffith et al., 1999). As noted in the literature review, these studies did not find significant correlations with vocabulary and the A not B task in their preschool samples either. In Nee et al.'s framework, an EF task based on spatial information would not be expected to be useful in improving our understanding of language development and therefore should not have been included in the test protocol of the current research. Further research is clearly needed to clarify models of the CE. In any case, no matter which interpretation of these data is preferred, the A not B task was not informative in predicting language outcomes in two to four year old children.

8.4.3 Executive functions (EF)

8.4.3.1 Behaviour Rating Inventory of Executive Functions – Preschool Version (BRIEF-P)

The BRIEF-P was chosen as the measure of EF. The reasons behind this choice were practical as well as theoretical, as outlined in the literature review. However this choice has led to some difficulties when it comes to interpreting the results. Firstly, construct validity was an issue. Parent report of problem behaviours is a subjective judgement; being influenced by parental expectations. Some parents commented that they were not sure what executive control behaviours could be reasonably expected of their two year old. In addition, problem behaviours in general can arise from a variety of causes (such as poor communication, emotional stressors and parenting style mismatch with child temperament). In other words there is a difference between a propensity towards difficult EF behaviours at home and capability in EF. Therefore these EF measures cannot be assumed to reflect difficulties with the CE. The results could be seen to be in line with these observations. Throughout Studies 1 and 2, the EFs correlated strongly with each other, but showed at most a moderate association with other measures, a trend particularly observable at ages 24-30 months. All five EFs in the BRIEF-P are measured in a broad behavioural sense, and as such may bear little relation to specific executive cognitive processes by the same names (Burgess et al., 2006), which are thought to underlie language acquisition. On the other hand, language itself would logically mediate the development of these executive behaviours. As a starting point for the following sections, I first assume each subscale measures what it purports to and then discuss construct validity and the likely causal directions between EF and language for each one. Throughout these sections I will refer to the two studies (Gioia et al., 2003; Wittke et al., 2013) which also investigated preschoolers with language difficulties using the BRIEF-P as comparison studies. Both studies reported generally lower EF (Global Executive Composite (GEC) scores) in children with LI than controls matched by age, gender and mother's education level. However there were some differences in the results between these studies and the current research which will be discussed below. Unfortunately a detailed comparison is not possible due to gaps in the reporting, for example, Gioia et al. (2003) did not report the criteria used for defining their LI group and Wittke et al. (2013) did not report scores by group for each of the five BRIEF-P subscales independently, but instead combined them into three indices.

8.4.3.2 Shift

Shift was the only EF (measured by the BRIEF-P) that contributed unique variance to expressive vocabulary / language at both time points tested and also longitudinally. These results were as expected and are in line with previous research, which showed an association between shifting and LI in school and preschool aged children (Gioia et al. (2003); Im-Bolter et al., 2006; Wittke et al., 2013). While the hypothesis was met, the interpretation of these data is debatable. One interpretation of these results is that poor shifting is a factor in early language acquisition. This would presumably be because an increased ability to shift from one mental scheme to another may allow for faster maturational progress in language. Recently, poor shifting (measured as cognitive flexibility in rule abstraction) has been association with phonological disorders (Dodd & McIntosh, 2010). If there is an effect of poor cognitive flexibility in learning phonological rules, similar effects could be expected in learning standard parameters for grammar, semantic categories or even pragmatics. However there is another possible interpretation of the association between poor Shifting and concurrent expressive language in the current cohort. This is that poor expressive language results in difficulties with Shifting. It is logical that children with better language skills can be better prepared for changes in their environment through verbal discussion, and therefore find transitions easier to cope with. Alternatively, this could be seen as a spurious result. The pattern of results seen in Study 3 supports this conclusion. If Shifting was a contributing factor in language acquisition, it would be expected that it would have an effect on future language (Study 3) in the same direction as it did with concurrent language skills (Studies 1 and 2). However in Study 3, more problems with Shifting at ages 24-30 months predicted better language at 41-49 months, whereas more problems with Shifting predicted worse concurrent language scores at both time points tested (Studies 1 and 2). The changing nature of the relationship between Shift and language between the two concurrent Studies (1 and 2) and the longitudinal one (Study 3) suggests a conservative approach to interpreting these data is appropriate.

In further support of cautious interpretation of these results, is is uncertain exactly what type of "shifting" is being measured. Several of the items used in this subscale could more closely correspond with measurement of a withdrawn or introverted personality rather than a difficulty with shifting between mental sets e.g. "has trouble adjusting to new people" and "has trouble joining in at unfamiliar social events". Several items also seem to measure sensory sensitivities rather than shifting e.g "is bothered by loud noises, bright lights or certain smells" and "acts overwhelmed or overstimulated in crowded busy situations". Therefore the degree to which this subscale actually measures the microlevel shifting hypothesised to be involved in language development is uncertain.

Overall, the pattern of evidence here is not convincing evidence of the hypothesis that poor EF constrains language progress. This was also the conclusion of Im-Bolter et al. (2006) from their path analysis study of EF in primary school aged children with LI.

8.4.3.3 Emotional Control

Emotional Control was the only other EF at 24-30 months to contribute unique variance in the regression analysis predicting total language outcomes. Children whose parents rated them as having more problem behaviours in Emotional Control at age 24-30 months, had worse language outcomes 18 months later. Emotional Control did not predict unique variance in concurrent expressive vocabulary / language at either age range tested, however it significantly correlated on a bivariate level. One interpretation of this is to take the Emotional Control ratings (alongside VWM and Shift) as evidence of problems with the CE, leading to poorer language outcomes over time. However Emotional Control is a meta-cognitive EF, rather than a simple aspect of cognitive control, such as inhibition. The Emotional Control subscale could just as easily be seen as an index of externalising frustration associated with struggling to communicate, as being an index of poor CE function. In support of this interpretation, a significant positive moderate correlation was seen between parent concern about their child's communication at ages 24-30 months (measured in the Parent Questionnaire I) and Emotional Control in Study 1, $r(73) = .38$, $p = .001$. Early parent concern regarding communication was also negatively correlated with total language scores at ages 41-49 months, $r\ (74) = -.44$, $p < .001$. Children develop emotional regulation skills partly by talking through how they feel with their parents in a responsive loving relationship. Poor communication would naturally slow this process. Other studies have found that once children's language difficulties resolved, their behavioural and social-emotional difficulties also resolved, leading to the conclusion that emotional disturbances in children with LI are a result of poor communication rather than part of the underlying difficulty (Fujiki et al., 2002; Irwin et al., 2002; Redmond & Rice, 1998). Overall the better interpretation seems to be that those children who had poorer language outcomes had more difficulties with communication and Emotional Control at age two years, presumably due to more frustration over their compromised language learning system.

8.4.3.4 Inhibition

Two previous studies found equivocal results on this measure. Using the BRIEF-P, Wittke et al. (2013) found no group differences in Inhibition in their group of pre-schoolers with SLI compared with TD controls (this information is inferred rather than explicitly reported in the text). Whereas Gioia et al. (2003) reported that 32% of their preschool children with LI had clinically significant difficulties with Inhibition. This indicates that children with LI are heterogeneous with regards to Inhibition as measured by the BRIEF-P. Previous studies of TD preschoolers have also found associations between response inhibition and language (Epsy et al., 1999; Viterbori et al., 2012; Willoughby et al., 2010). In the current research the only significant associations between language and Inhibition were small in size and were found between the PLS-4 AC scores and PLS-4 total language scores and Inhibition at age 41-49 months only. There were no significant group differences in Inhibition between late talkers and TD children at 24-30 months or between the RLT and TD groups at 41-49 months. Overall the hypothesis for a role of response inhibition in early language acquisition was not supported by these data. The few significant associations found between response inhibition and language, in this and prior research, are easier to account for as being causal in the other direction. Better language skills enable children to understand explanations of how they need to behave, and this presumably helps them to develop impulse control. Alternatively, as for the whole of the BRIEF-P assessment, issues with the measure itself may be responsible for the null findings.

8.4.3.5 Working Memory

Despite the fact that the BRIEF-P was developed for two to five year olds, there were concerns about the developmental appropriateness of some of the Working Memory items. For example, “unable to finish describing an event person or story” (Item 55) is not suitable for the 10-20% of two year old children still speaking in single words. Even some non-linguistic items were difficult for parents to apply to their toddlers (e.g. Item 53 “does not try as hard as his or her ability on activities”) is a difficult statement to apply to the free-play learning environment of two year olds. Anecdotally I noted parents circled “never a problem” for items describing skills their child had not yet developed. Developmental inappropriateness of certain items affected every BRIEF-P subscale (but particularly Working Memory and Plan / Organise). The level of difficulty with working memory function in everyday life may therefore have been underestimated in the 24-30 month olds. By 41-49 months, this problem of developmental inappropriateness had resolved for all but the children with severe language delays. In support of these observations, the KWM I task and Working Memory did not correlate at ages 24-30 months, but correlated to a moderate degree at ages 41-49 months.

Despite these difficulties, the Working Memory subscale showed significant moderate bivariate correlations with language at 24-30 months and 41-49 months. There were significant group differences in Working Memory scores between late talkers and TD children at 24-30 months, and the RLT and TD groups at 41-49 months. It is possible if the items were more developmentally appropriate to the age range tested here, that stronger associations between Working Memory and language may have been seen. Wittke et al. (2013) and Gioia et al. (2003) also reported lower Working Memory scores in their pre-schoolers with LI compared with controls, indicating a stronger trend for Working Memory and language to be associated than other EFs. Working Memory was not a significant predictor for expressive vocabulary (at 24-30 months) and expressive language (at 41-49 months) on a multivariate level. This is not surprising given the overlap in constructs (e.g. with the KWM tasks). Overall the Working Memory subscale was largely redundant as a predictor in this research due to the success of the KWM task as a measure of VWM.

8.4.3.6 Plan / Organise

It was thought that better Planning / Organising ability might contribute to language acquisition by allowing better formulation of early discourse and narratives. The only significant association Plan / Organise showed with expressive language were the bivariate correlations at 41-49 months. The lack of similar correlations at ages 24-30 months may have been due to developmental inappropriateness of some items for this age range (as discussed for the Working Memory subscale). Being a macro-level EF which builds on several others, it is not surprising that Plan / Organise was no longer a significant predictor of concurrent expressive language once more specific measures of EF (e.g. VWM) was entered into the multivariate regression model. There were no significant group differences between the late talker and TD groups at 24-30 months or the RLT and TD groups at 41-49 months, although the trend was in that direction. The two other studies using the BRIEF-P with preschoolers with LI reported equivocal findings (Gioia et al. (2003); Wittke et al., 2013). This is most likely due to heterogeneity in the populations tested, possibly combined with small sample sizes (*n* = 19 and 21). This current research does not provide any evidence to suggest that Plan / Organise is a factor in early language development or in late talking. A more likely interpretation of these results is that better language skills enable better planning and organisation development and that the BRIEF-P measures this more effectively at ages 41-49 months rather than 24-30 months.

8.4.3.7 Summary of Executive Functions (EF) and language

Only two EFs from the BRIEF-P played a unique role in predicting outcomes in language, and both were limited to small influences in multivariate models. Children with poorer early Emotional Control were more likely to have poorer language outcomes. Increased difficulties with Emotional Control may indicate more stress around communication in the home and a greater underlying dysfunction with language learning. Poorer Shifting was associated with worse expressive language at both age ranges tested in both bivariate and multivariate models. The most likely explanation of this is that better language skills facilitate better social shifting through verbal discussion and reassurance. Cautious interpretation of these data is warranted as poor early Shifting predicted better language outcomes in this cohort, which is likely to have been a spurious result. With the possible exception of Working Memory (which was superseded by the presence of two other working memory measures in the test protocol), none of the EFs measured by the BRIEF-P provided strong evidence for CE deficits being involved in late talking and or language development. In contrast, the associations between EF and language are more likely to reflect the supportive role language ability plays in development of EF behaviours than vice versa. To conclude, the BRIEF-P is unlikely to be a useful addition to a test protocol aiming to predict language outcomes. However it does provide information about parent perception of the child's behaviour, which may influence a clinical decision to intervene immediately to reduce stress in the household, regardless of the child's language prognosis.

As a final comment, the BRIEF-P model is organised by function, in accordance with an assumption of the domain generality of the CE (Baddeley, 2007). The interpretations above are made according to this model. The BRIEF-P results cannot be reinterpreted according to Nee et al.'s model (2013) (which assumes division of EF by content type), as most items on the questionnaire describe executive behaviours which cross domains. If Nee et al.'s model is supported by further research, many EF assessments and research will need to be reconsidered, including the BRIEF-P.

8.5 Models of individual prediction

The high degree of resolution of early expressive language delay (83%) by ages 41-49 months in this cohort was unexpected. Other late talker studies have reported lower rates of resolution by age four years. For example, Paul (1993) reported 47%; Rescorla, Dahlsgaard, and Roberts (2000) reported 71% and Dale et al. (2003) reported 40%. There are several differences between these studies and the current research, including the criteria and ages used to identify late talkers, the cut-off for LI and the types of measures used. On the basis of these differences, I had hypothesised that the resolution rate in this study would be 50-75%, but this figure was an underestimate. The use of the PLS-4 as the outcome measure may have inflated resolution rates. Multiphasic measures of language may underestimate those having specific difficulties in morphosyntax; which is known to lag behind in LI. For example, Rescorla et al. (2000) reported a 71% rate of resolution by age four years using the 10th percentile on mean length of utterance, but this fell to 29% in the same cohort when using the 10th percentile on measure of syntax (IP-Syn; Scarborough (1990)). The distribution of participants in our current research was also skewed towards the high end of educational achievement as shown in Chapter 3. Environmental factors have a cumulative effect on language over time, whether positive or negative. Late talkers with reduced linguistic stimulation in the home are likely to be at more risk for ongoing delays than those in advantaged households (Reilly et al., 2010). The high educational level of the parents of many of the current cohort may also have inflated the rates of resolution (although the middle class was noted to be over-represented in the studies cited above also).

Dollaghan (2013) made a case for calling for a moratorium on the use of "late talking" as a clinical category for identification of children in need of services, on the basis that the characteristics of late talking do not meet the requirements for a screening programme (Wilson and Junger,1968 as cited in Dollaghan, 2013). The results of the current research do not strongly support this conclusion. In this cohort, 17% of late talkers met the criteria for LI at age 41-49 months, compared with only 2% of the TD group. Late talkers, even in this high functioning cohort, were seven times more likely to be monitored or referred to a speech language therapist at follow up in this study, than TD children (see Study 3). In addition, two of the late talkers were diagnosed with autism before the outcome assessments were completed. This process was facilitated by identifying these children early as late talkers. However, most of the clinical concerns registered at 41-49 months related to phonological delays, which are likely to resolve with time (Dodd et al., 2002). Phonological delays have also been reported in other late talker cohorts (Paul, 1993; Roberts, Rescorla, Giroux, & Stevens, 1998). Perhaps the category of late talker should be considered as one of several risk factors rather than a sole identifier of children at risk (such as the approach of Reilly et al. (2010) or the recommendations of Whitehurst and Fischel (1994)). Late talking remains a useful red flag however, as parents can readily identify late talking but may miss signs of other delays in development. Late talking also continues to be a useful category for research purposes, as a comparison group to LI and TD children, as is demonstrated by the current research.

The hypothesis that adding working memory variables to existing predictive models for language outcomes would improve individual prediction was not upheld. In fact the best predictive model for individuals in this study comprised a single factor: low early receptive language. Several literature reviews on this topic have also suggested receptive language is a moderate predictor of language outcomes in the preschool years (Ellis & Thal, 2008; Paul & Roth, 2011; Rescorla, 2011). However, early receptive language delays did not predict all cases of later LI on an individual level. One child, who had an expressive SLI at 41-49 months, was not low in receptive language at intake. Other children with early receptive delays did resolve over the next 18 months (4/9). Presumably VWM did not improve prediction of individual outcomes because of the near complete overlap between early low receptive language and low VWM. As all late talkers had poor PSTM initially, this variable did not distinguish between those who resolved and those who did not. Finally, the EF and processing speed measures showed too much variability in late talkers to be useful as predictors for individual outcomes. Bishop (2006) noted the role of compensation in children who have only one processing weakness. This could have been the mechanism at work in the RLTs, who remained low in PSTM as a group, but were scoring above average in language by ages 41-49 months. Possible compensatory factors (which were not measured in this cohort) might have been statistical learning skills or better quality and quantity of linguistic input from the environment. These factors could be explored in future research.

8.6 Limitations of the current research

One limitation of this study was the sample size. A larger sample size would have resulted in a greater number of children with receptive language delays at intake and children with LI at follow–up. Low numbers in these two key groups meant that statistical analyses could only be used to make robust group comparisons between late talkers, resolved late talkers and TD children. The results from the wider Learning to Talk study (*n* = 168) should allow for further analyses of these groups for some constructs (such as, PSTM). However, measures of the CE (VWM, processing speed and EF) were not included for the full cohort.

It was not possible for the outcome assessments to be done blind to initial group status. Therefore examiner bias in scoring cannot be ruled out. The sample was also skewed towards the higher end of parent education levels. All but one of the late talkers in the current study, whose sole difficulty at intake was expressive vocabulary, resolved. This child's mother had no formal qualifications and was observed to speak less frequently (both to her child and the researcher) than other mothers. Perhaps if there had been a wider spread of parental education in the cohort, this factor (parental education) may have been identified as a predictor of language outcomes over time. Finally the LWL II task seemed to elicit a more complex response from the children at ages 41-49 months compared with the simple shift in eye gaze registered at age two years, meaning comparisons in the role of processing speed across the two age ranges measured could not be made. An alternative processing speed task for ages 41-49 months is required should this study design be replicated.

8.7 Strengths of the current research

The longitudinal nature of the design allowed for the likely direction of the relationship between working memory variables and language to be explored. The lengthy test protocol meant a wide perspective on each child's functioning could be taken. The face-to-face nature of most of the assessment data was also a strength, as opposed to the heavy reliance some studies make on parent report. My clinical experience in gaining compliance from young children meant there was a low number of missing data points for the behavioural measures. The commitment of the parents to the study contributed to an exceptionally low rate of attrition over time (one child), which meant the outcomes of all the late talkers could be tracked.

8.8 Implications for clinical practise

The original question which prompted this research was: "Is there anything wrong with my late talking toddler?" This study has identified poor PSTM as a possible cause of transient delays in expressive vocabulary for some children, while acknowledging that further proofs are required. Interventions for toddlers with emerging language typically involve increasing exposure to language forms within the child's zone of proximal development (Paul, 2007). This should facilitate learning in children with early expressive language delays, by providing immediate compensation for their assumed poor PSTM skills.

The problem of distinguishing between transient and persistent delays remains unresolved; although a high degree of resolution (83%) was seen in the current cohort, and early receptive language delay was a stronger risk factor for poor language outcomes than late talking. Unfortunately, none of the working memory variables measured in this study improved prediction of individual outcomes over time, despite the fact that early VWM, Shift and Emotional Control contributed unique variance in total language outcomes on a group level. Therefore, at this stage, these measures do not warrant inclusion in clinical testing batteries for language delayed toddlers. The multifactorial risk factor models put forward by Ellis and Thal (2008), Paul and Roth (2011), Desmarais et al. (2008) and Rescorla (2011) (as summarised in the literature review) continue to represent some of the best available practical advice on late talker management for clinicians. However their models continue to be untested as a whole, and await confirmation of their diagnostic accuracy. In the interim, monitoring is a practical solution to any uncertainty about a child's prognosis.

8.9 Future research

This study has prompted questions around the exact nature of the relationship between working memory and early language acquisition. Several future avenues for research have already been mentioned throughout this discussion chapter. I would also like to suggest that further work be done exploring the nature of the interactions between VWM and language development in the preschool years. A randomised controlled study investigating the effect of VWM training compared with receptive language intervention in children with early language delays would increase our understanding of this area.

8.10 Conclusion

There is much debate in the literature regarding the main direction of influence between working memory and language development at different ages. Working memory and language develop as a dynamic integrated neural network and while they can be distinguished, they cannot be fully separated. The results of the current research could be seen to support the concept that capacity limits in working memory may constrain the development of certain language skills. For the first time, PSTM was included in a prospective study predicting language outcomes in TD and late talking children over the age range 24-49 months. Results indicated that PSTM is strongly involved in early expressive vocabulary acquisition, to the extent of appearing to be a causal factor of late talking in this cohort. Processing speed did not add unique variance in language scores and results suggest working memory mediates the relationship between processing speed and language ability in this age range. This study has also successfully measured VWM in two year old children. The results suggested a unique role for VWM in early language acquisition, particularly over time. Contrary to expectations, EF as measured by parent report was largely unhelpful in predicting language and in answering the question of whether deficits in the CE impact on language development. Overall there was more evidence in support of early language supporting EF than vice versa. This interpretation is made cautiously, as questions were raised about construct validity and the suitability of the BRIEF-P for measuring EF in late talkers. Evidence has been presented that domain general models of the CE may be inaccurate and if true, this calls into question the validity of the BRIEF-P subscales as measures of CE function.

To conclude, this cohort had a high rate of resolution of late talking, possibly due to the limited spread of SES status in the sample. However this lack of variety in environmental factors may have contributed to the strength of findings in the processing measures' relationship with language. None of the working memory variables here improved prediction of individual outcomes, instead, low early receptive language scores emerged as the best predictor of later LI for individuals in this cohort. In conclusion while this study has made a positive contribution to the knowledge base on working memory and language acquisition, a strong predictive model for language outcomes in individual children remains elusive.

APPENDIX A: Ethics Approval

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

HUMAN ETHICS COMMITTEE

Secretary, Lynda Griffioen
Email: human-ethics@canterbury.ac.nz

Ref: HEC 2011/121

1 December 2011

Professor Thomas Klee
Department of Communication Disorders
UNIVERSITY OF CANTERBURY

Dear Thomas

The Human Ethics Committee advises that your research proposal "Early factors in childhood communication disorders" has been considered and approved.

Please note that this approval is subject to the incorporation of the amendments you have provided in your email of 1 December 2011.

Best wishes for your project.

Yours sincerely

Michael Grimshaw
Chair
University of Canterbury Human Ethics Committee

University of Canterbury Private Bag 4800, Christchurch 8140, New Zealand. www.canterbury.ac.nz

APPENDIX B: Parent Pack for Part 1

Learning to Talk
Part 1

Information for parents/whānau

An invitation to participate in part 1 of a research project on children's early language development

Kia ora! Hello!
We would like to invite you to take part in a research project about children's early language development. Before deciding if you'd like to participate, please read through this information. If you have any questions about the project, feel free to get in touch with us by phone or email.

What is the purpose of the project?
Some children learn to talk quickly and some take their time. Many two-year-olds are joining words together in sentences while others are saying only a few words. Some children who start off slowly catch up over time, while others have ongoing problems with language. The purpose of this study is to explore the range of language abilities of young children in New Zealand and improve the way in which those with speech and language difficulties are identified.

What is involved?
The project will take place in two parts. In the first part, we are searching for children between 24 and 30 months of age. If your child will be in this age range between now and December 2012, you can participate! You need to fill in two questionnaires. One questionnaire is about your child's use of words and sentences. The other is about your family and your child's birth history. We need 1000 parents in the Canterbury region to complete these two questionnaires. Whether your child is talking a lot or has not yet begun to talk, we would like to hear from you.

In the second part of the project, we will invite around 200 of you to bring your child to the University of Canterbury, where a speech and language therapist will assess your child's speech, language, hearing and memory skills. There is another information sheet about that part of the study which you can request when you send back the questionnaires for part 1.

Even if you don't think you can commit the time to the second part of the project, we would appreciate it if you would fill out the two questionnaires needed for the first part of the project. This should take no more than 20 to 30 minutes and your involvement will end there.

What will happen to the questionnaires I complete?
None of the personal information you provide us will be made public. Your responses on the questionnaires will be kept confidential. They will be combined with those of other participants and summarised. Your questionnaires will be kept in a secure locked cabinet at the University. Only people working on the project will have access to them. At the end of the project, the questionnaires will be destroyed.

We will write up the results of the research and share them with people who work with children across the world. Summaries of our research may be presented at conferences, published in scientific and professional journals, or appear on our website. Jayne Moyle will also write up her part of the project as a doctoral thesis, which will be available in the University of Canterbury library. If you would like to receive a brief summary of our findings, you can tick the box on the consent form and we will send you one when the study is completed.

How can I volunteer to take part?
If you would like to take part in the first part of this project, please contact us (contact details below) and we will send you the questionnaires, consent form and a postage-free envelope to return them to us once completed.

If you would like to take part in the second phase of this project, you can tick the box at the bottom of the consent form and we will contact you to let you know if we need you further.

If you have any concerns about your child's speech and language development, you can talk about them with us.

There aren't any risks to you or your child as a result of participating in this study.

You are able to pull out of the study at any time and you don't have to give a reason.

Other information:
This research project is funded by the Marsden Fund of the Royal Society of New Zealand (research project no. UOC1003, 2011-2014). The project has been reviewed and approved by the University of Canterbury Human Ethics Committee (HEC 2011/121) and the Plunket Ethics Committee.

The research team: Professor Thomas Klee, Professor Stephanie Stokes, Dr Catherine Moran, Jayne Moyle, Brynlea Collin, Christine Shalders and Louise Hughes
Address: Child Language Centre | Te Reo o te Tamaiti
7 Creyke Road, Ilam, Christchurch 8041
Phone: 364 2987 ext. 6431 or 8193. Cell phone: 021 0282 7225
Email: ChildLanguageCentre@canterbury.ac.nz
Website: www.cmds.canterbury.ac.nz/clc/

Consent Form

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Learning to Talk - Part 1

I have read and understood the information that was given to me about the research project named above. I have had an opportunity to ask questions and have had them answered.
I understand that my participation in this project is voluntary. I understand that the information you collect from me will remain confidential and will be stored securely at the university.
I understand that any presentations or publications resulting from this project will not refer to me or anyone in my family by name. I understand that this project has been reviewed ***and approved*** by the University of Canterbury Human Ethics Committee. On this basis, I agree to participate in this research project.

MY CHILD'S NAME (please print):...

My NAME (please print): ....

My signature: ....

Date: ....

Address: ....

....

❑ Feel free to contact me about participating in part 2 of this project. I understand that I am under no obligation to participate.
❑ I would like you to send me a brief summary of your findings when the study is complete.

Please contact me by:

- ❑ Email ____
- ❑ Cell phone ____
- ❑ Landline ____
- ❑ Post

Main Researcher:
Professor Thomas Klee
University of Canterbury
Private Bag 4800, Christchurch 8140
Email: ChildLanguageCentre@canterbury.ac.nz
Phone: 03 364 2987 ext. 8501

Child's Name ______________________________	**Sex** __________
Birthdate ______________ **Age** __________*	**Today's Date** ______________

** Please check your child is aged 2-2.5 years old at the time you fill in this questionnaire*

The MacArthur-Bates Communicative Development Inventory: Words & Sentences

New Zealand English Adaptation

PART 1 – WORDS CHILDREN USE

A. VOCABULARY CHECKLIST

Children understand many more words than they say. We are particularly interested in the words your child SAYS. Please go through the list and mark the words you have heard your child use. If your child uses a different pronunciation of a word (for example, "raffe" instead of "giraffe" or "sketti" for "spaghetti"), mark the word anyway. Remember that this is a "catalogue" of all the words that are used by many different children. Don't worry if your child only knows a few of these right now.

1. SOUND EFFECTS AND ANIMAL SOUNDS (12)

baa baa	O	meow	O	uh oh	O
choo choo	O	moo	O	vroom	O
cockadoodledoo	O	ouch	O	woof woof	O
grr	O	quack quack	O	yum yum	O

2. ANIMALS (Real or Toy) (42)

alligator	O	duck	O	owl	O
animal	O	elephant	O	penguin	O
ant	O	fish	O	pig	O
bear	O	frog	O	pony	O
bee	O	giraffe	O	possum	O
bird	O	goose	O	puppy	O
bunny	O	hen	O	rooster	O
butterfly	O	horse	O	sheep	O
cat	O	insect	O	teddybear	O
chicken	O	lamb	O	tiger	O
cow	O	lion	O	turkey	O
deer	O	monkey	O	turtle	O
dog	O	moose	O	wolf	O
donkey	O	mouse	O	zebra	O

3. VEHICLES (Real or Toy) (14)					
aeroplane	O	fire engine	O	tractor	O
bike	O	helicopter	O	train	O
boat	O	motor bike	O	tricycle	O
bus	O	pushchair*	O	truck	O
car	O	sled	O		

*or word used in your family: please add to Section F.

4. TOYS (18)					
ball	O	chalk	O	pencil	O
balloon	O	crayon	O	play dough	O
bat	O	doll	O	present	O
block	O	game	O	puzzle	O
book	O	glue	O	story	O
bubbles	O	pen	O	toy	O

5. FOOD AND DRINK (68)					
apple	O	fizzy drink	O	peas	O
banana	O	food	O	pizza	O
beans	O	gherkin	O	popcorn	O
biscuit	O	grapes	O	potato	O
bread	O	green beans	O	potato chip	O
butter	O	hamburger	O	pretzel	O
cake	O	ice	O	pudding	O
carrots	O	ice block	O	pumpkin	O
cereal	O	ice cream	O	raisin	O
cheerios	O	jam	O	salt	O
cheese	O	jelly	O	sandwich	O
chewing gum	O	juice	O	sauce	O
chicken	O	lollies	O	soup	O
chips	O	lollipop	O	spaghetti	O
chocolate	O	meat	O	strawberry	O
coffee	O	melon	O	toast	O
coke	O	milk	O	tomato sauce	O
corn	O	muffin	O	tuna	O
cracker	O	noodles	O	vanilla	O
doughnut	O	nuts	O	vitamins	O
drink	O	orange	O	water	O
egg	O	pancake	O	yogurt	O
fish	O	peanut butter	O		

6. CLOTHING (27)					
beads	O	jacket	O	shorts	O
belt	O	jeans	O	slipper	O
bib	O	jersey	O	sneaker	O
boots	O	nappy	O	snowsuit	O
button	O	necklace	O	sock	O
coat	O	pyjamas	O	tights	O
dress	O	scarf	O	trousers	O
gloves	O	shirt	O	underpants	O
hat	O	shoe	O	zip	O

7. BODY PARTS (27)					
ankle	O	feet	O	nose	O
arm	O	finger	O	penis*	O
belly button	O	hair	O	shoulder	O
buttocks/bottom/bum*	O	hand	O	sore	O
cheek	O	head	O	tooth	O
chin	O	knee	O	toe	O
ear	O	leg	O	tongue	O
eye	O	lips	O	tummy	O
face	O	mouth	O	vagina*	O

*or word used in your family: please add to Section F.

8. SMALL HOUSEHOLD ITEMS (49)					
basket	O	glasses	O	purse	O
blanket	O	hammer	O	radio	O
bottle	O	jar	O	rubbish	O
box	O	keys	O	scissors	O
bowl	O	knife	O	serviette	O
broom	O	lamp	O	soap	O
brush	O	light	O	spoon	O
bucket	O	medicine	O	tape	O
camera	O	money	O	telephone	O
clock	O	mop	O	tin	O
coin	O	nail	O	tissue	O
comb	O	paper	O	toothbrush	O
cup	O	picture	O	towel	O
dish	O	pillow	O	tray	O
fork	O	plant	O	vacuum cleaner*	O
glass	O	plate	O	walker	O
				watch	O

*or word used in your family: please add to Section F.

9. FURNITURE AND ROOMS (33)					
bath	O	dryer	O	room	O
bathroom	O	fridge	O	shower	O
bed	O	garage	O	sink	O
bedroom	O	high chair	O	sofa	O
bench	O	kitchen	O	stairs	O
cellar	O	living room	O	stove	O
chair	O	oven	O	table	O
cot	O	playpen	O	TV	O
couch	O	porch	O	wardrobe	O
door	O	potty	O	washing machine	O
drawer	O	rocking chair	O	window	O

10. OUTSIDE THINGS (31)					
backyard	O	pool	O	star	O
cloud	O	rain	O	stick	O
flag	O	rock	O	stone	O
flower	O	roof	O	street	O
footpath	O	sandpit	O	sun	O
garden	O	sky	O	swing	O
grass	O	slide	O	tree	O
hose	O	snow	O	water	O
ladder	O	snowman	O	wind	O
lawn mower	O	spade	O		
moon	O	sprinkler	O		

11. PLACES TO GO (22)

beach	O	home	O	playground	O
camping	O	house	O	school	O
church*	O	movie	O	shop	O
circus	O	outside	O	work	O
country	O	park	O	yard	O
downtown	O	party	O	zoo	O
farm	O	petrol station	O		
forest	O	picnic	O		

*or word used in your family: please add to Section F.

12. PEOPLE (29)

aunt/auntie	O	daddy*	O	mummy*	O
baby	O	doctor	O	nurse	O
babysitter	O	fireman	O	people	O
babysitter's name	O	friend	O	person	O
boy	O	girl	O	pet's name	O
brother	O	grandma*	O	police	O
child	O	grandpa*	O	sister	O
child's own name	O	lady	O	teacher	O
clown	O	mailman	O	uncle	O
cowboy	O	man	O		

*or word used in your family: please add to Section F.

13. GAMES AND ROUTINES (24)

bath	O	hi	O	please	O
breakfast	O	hello	O	shh/shush/hush	O
bye	O	lunch	O	shopping	O
call (on the phone)	O	nap	O	snack	O
dinner/tea	O	night night	O	thank you	O
give me five!	O	no	O	this little piggy	O
gonna get you!	O	patty cake	O	turn around	O
go potty	O	peekaboo	O	yes	O

14. ACTION WORDS (103)

bite	O	drive	O	hug	O	read	O	swim	O
blow	O	drop	O	hurry	O	ride	O	swing	O
break	O	dry	O	jump	O	rip	O	take	O
bring	O	dump	O	kick	O	run	O	talk	O
build	O	eat	O	kiss	O	say	O	taste	O
bump	O	fall	O	knock	O	see	O	tear	O
buy	O	feed	O	lick	O	shake	O	think	O
carry	O	find	O	like	O	share	O	throw	O
catch	O	finish	O	listen	O	show	O	tickle	O
chase	O	fit	O	look	O	sing	O	touch	O
clap	O	fix	O	love	O	sit	O	wait	O
clean	O	get	O	make	O	skate	O	wake	O
climb	O	give	O	open	O	sleep	O	walk	O
close	O	go	O	paint	O	slide	O	wash	O
cook	O	hate	O	pick	O	smile	O	watch	O
cover	O	have	O	play	O	spill	O	wipe	O
cry	O	hear	O	pour	O	splash	O	wish	O
cut	O	help	O	pretend	O	stand	O	work	O
dance	O	hide	O	pull	O	stay	O	write	O
draw	O	hit	O	push	O	stop	O		
drink	O	hold	O	put	O	sweep	O		

15. DESCRIPTIVE WORDS (63)

allgone	O	full	O	orange	O
asleep	O	gentle	O	poor	O
awake	O	good	O	pretty	O
bad	O	green	O	quiet	O
better	O	happy	O	red	O
big	O	hard	O	sad	O
black	O	heavy	O	scared	O
blue	O	high	O	sick	O
broken	O	hot	O	sleepy	O
brown	O	hungry	O	slow	O
careful	O	hurt	O	soft	O
clean	O	last	O	sticky	O
cold	O	little	O	stuck	O
cute	O	long	O	thirsty	O
dark	O	loud	O	tiny	O
dirty	O	mad	O	tired	O
dry	O	naughty	O	wet	O
empty	O	new	O	white	O
fast	O	nice	O	windy	O
fine	O	noisy	O	yellow	O
first	O	old	O	yucky	O

16. WORDS ABOUT TIME (12)

after	O	morning	O	today	O
before	O	night	O	tomorrow	O
day	O	now	O	tonight	O
later	O	time	O	yesterday	O

17. PRONOUNS (25)

he	O	me	O	their	O	we	O
her	O	mine	O	them	O	you	O
hers	O	my	O	these	O	your	O
him	O	myself	O	they	O	yourself	O
his	O	our	O	this	O		
I	O	she	O	those	O		
it	O	that	O	us	O		

18. QUESTION WORDS (7)

how	O	when	O	which	O	why	O
what	O	where	O	who	O		

19. PREPOSITIONS AND LOCATIONS (26)

about	O	down	O	on top of	O
above	O	for	O	out	O
around	O	here	O	over	O
at	O	inside/in	O	there	O
away	O	into	O	to	O
back	O	next to	O	under	O
behind	O	of	O	up	O
beside	O	off	O	with	O
by	O	on	O		

20. QUANTIFIERS AND ARTICLES (17)

a	O	each	O	other	O
all	O	every	O	same	O
a lot	O	more	O	some	O
an	O	much	O	the	O
another	O	none	O	too	O
any	O	not	O		

21. HELPING VERBS (21)

am	O	does	O	need/ need to	O
are	O	don't	O	try/ try to	O
be	O	gonna/ going to	O	wanna/ want to	O
can	O	gotta/ got to	O	was	O
could	O	hafta/ have to	O	were	O
did/ did ya	O	is	O	will	O
do	O	lemme/ let me	O	would	O

22. CONNECTING WORDS (6)

and	O	but	O	so	O
because	O	if	O	then	O

B. HOW CHILDREN USE WORDS	Not Yet	Sometimes	Often
1. Does your child ever talk about past events or people who are not present? For example, a child who saw a parade last week might later say parade, clown or band.	O	O	O
2. Does your child ever talk about something that's going to happen in the future, for example, saying "choo choo" or "aeroplane" before you leave the house for a trip, or saying "swing" when you are going to the park?	O	O	O
3. Does your child talk about objects that are not present such as asking about a missing or absent toy, referring to a pet out of view, or asking about someone not present?	O	O	O
4. Does your child understand if you ask for something that is not in the room, for example, by going to the bedroom to get a teddy bear when you say "where's the bear?"	O	O	O
5. Does your child ever pick up or point to an object and name an absent person to whom the object belongs? For example, a child might point to mummy's shoe and say "mummy".	O	O	O

PART II – SENTENCES AND GRAMMAR

A. WORD ENDINGS/PART I	Not Yet	Sometimes	Often
1. To talk about more than one thing, we add an 's' to many words. Examples include cars (for more than one car), shoes, dogs and keys. Has your child begun to do this?	O	O	O
2. To talk about ownership, we add an "'s", for example, Daddy's key, cat's dish and baby's bottle. Has your child begun to do this?	O	O	O
3. To talk about activities, we sometimes add 'ing' to verbs. Examples include looking, running and crying. Has your child begun to do this?	O	O	O
4. To talk about things that happened in the past, we often add 'ed' to the verb. Examples include kissed, opened and pushed. Has your child begun to do this?	O	O	O

B. WORD FORMS

Following are some other words children learn. Please mark any of these words that your child uses.

NOUNS

children	O	men	O	teeth	O
feet	O	mice	O		

VERBS

ate	O	fell	O	made	O
blew	O	flew	O	ran	O
bought	O	got	O	sat	O
broke	O	had	O	saw	O
came	O	heard	O	took	O
drank	O	held	O	went	O
drove	O	lost	O		

C. WORD ENDINGS/PART 2

Young children often place the wrong endings on words. For example, a child might say "Auntie goed home". Mistakes like this are often a sign of progress in language. In the following lists, please mark all the mistakes of this kind you have heard your child say recently.

NOUNS

blockses	O	mans	O	sockses	O
childrens	O	mens	O	teeths	O
childs	O	mices	O	toeses	O
feets	O	mouses	O	tooths	O
foots	O	shoeses	O		

VERBS

ated	O	comed	O	goed	O	ranned	O
blewed	O	doed	O	gotted	O	runned	O
blowed	O	dranked	O	haved	O	seed	O
bringed	O	drinked	O	heared	O	satted	O
buyed	O	eated	O	holded	O	sitted	O
breaked	O	falled	O	losed	O	taked	O
broked	O	flied	O	losted	O	wented	O
camed	O	getted	O	maked	O		

HAS YOUR CHILD BEGUN TO COMBINE WORDS YET, SUCH AS "NOTHER CRACKER", OR "DOGGIE BITE"?

O Not Yet O Sometimes O Often

IF YOU ANSWERED NOT YET, PLEASE STOP HERE. IF YOU ANSWERED SOMETIMES OR OFTEN, PLEASE CONTINUE.

D. EXAMPLES: Please list three of the longest sentences you have heard your child say recently.

1. ____________________

2. ____________________

3. ____________________

E. COMPLEXITY

In each of the following pairs, please mark the one that sounds MOST like the way your child talks right now. If your child is saying sentences even longer or more complicated than the two provided, just pick the second one.

1. Two shoe. O
   Two shoes. O
2. Two foot. O
   Two feet. O
3. Daddy car. O
   Daddy's car. O
4. (Talking about something happening right now.)
   Cat sleep. O
   Cat sleeping. O
5. (Talking about something happening right now.)
   I make tower. O
   I making tower. O
6. (Talking about something that already happened.)
   I fall down. O
   I fell down. O
7. More biscuit! O
   More biscuits! O
8. These my tooth. O
   These my teeth. O
9. Baby blanket. O
   Baby's blanket. O
10. (Talking about something that already happened.)
    Doggie kiss me. O
    Doggie kissed me. O
11. (Talking about something that already happened.)
    Daddy pick me up. O
    Daddy picked me up. O
12. (Talking about something that already happened.)
    Doggie go away. O
    Doggie went away. O
13. Doggie table. O
    Doggie on table. O
14. That my truck. O
    That's my truck. O
15. Baby crying. O
    Baby is crying. O
16. You fix it? O
    Can you fix it? O
17. Read me story, Mummy. O
    Read me a story, Mummy. O
18. No wash dolly. O
    Don't wash dolly. O
19. Want more juice. O
    Want juice in there. O
20. There a cat. O
    There's a cat. O
21. Go bye-bye. O
    Wanna go bye-bye. O
22. Where Mummy go? O
    Where did Mummy go? O
23. Coffee hot. O
    That coffee hot. O
24. I no do it. O
    I can't do it. O
25. I like read stories. O
    I like to read stories. O
26. Don't read book. O
    Don't want you read that book. O
27. Turn on light. O
    Turn on the light so I can see. O
28. I want that. O
    I want that one you got. O
29. Want biscuits. O
    Want biscuits and milk. O
30. Biscuit Mummy. O
    Biscuit for Mummy. O
31. Baby want eat. O
    Baby want to eat. O
32. Lookit me! O
    Lookin me dancing! O
33. Lookit! O
    Lookit what I got! O
34. Where's my dolly? O
    Where's my dolly name Sam? O
35. We made this. O
    Me and Paul made this. O
36. I sing song. O
    I sing song for you. O
37. Baby crying. O
    Baby crying cuz she's sad. O

F. OTHER COMMENTS/ other words your child says:

THANK YOU FOR COMPLETING THIS.

Parent Questionnaire

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Please complete this questionnaire and the *Communicative Development Inventory: Words and Sentences* and return both in the enclosed envelope. If you'd rather talk to us over the phone instead of filling out written questionnaires, just let us know!

1. Child's name: ___________________________________
2. Child's birth date: Day _____ Month _____ Year 20_____
3. Child's gender: Male _____ Female _____
4. Child's birth order: 1st _____ 2nd _____ 3rd _____ 4th or more _____
5. Child's birth weight: __________ grams (*or* _____lbs _____oz)
6. Was this child born a twin? No _____ Yes _____
7. Was this child born prematurely? No _____ Yes _____
   ↳ If yes, by how many weeks? _____
8. Has this child ever had any major health problems?
   No _____ Yes _____
   ↳ If yes, what are they? ____
9. Number of children in family, *including* this child: _____
10. In which country was this child born?
    ☐ New Zealand
    ☐ other; please indicate where: ______________________________________
    ↳ If other, how long has this child lived in New Zealand? _____________________
11. Which ethnic group does this child belong to? Tick the one or ones which apply.
    ☐ New Zealand European
    ☐ Māori
    ☐ Samoan
    ☐ Cook Island Maori
    ☐ Tongan
    ☐ Niuean
    ☐ Chinese
    ☐ Indian
    ☐ other such as Dutch, Japanese, Tokelauan; please state: ______________________
12. What is this child's main language? ____
13. Are any other languages spoken in the child's home?
    No _____ Yes _____
    ↳ If yes, which ones? ____

Please continue on the next page.

14. Is this child in day care or an early childhood education programme or cared for regularly by anyone else?
No _____ Yes _____
↳ If yes, how many hours per week on average? _____

15. Do you have any concerns about this child's ability to hear? Yes ___ No ___

16. Do you have any concerns about this child's language development? Yes ___ No ___

17. Do you have any concerns about this child's ability to communicate? Yes ___ No ___

18. If you answered "Yes" to any of the last 3 questions, say why you are concerned:

____

____

____

19. Would you like us to get in touch with you to discuss your concerns? Yes ___ No ___

☐ Contact me on my home phone at: ________________________________

☐ Contact me on my cell phone at: ________________________________

☐ Contact me by email post at: ________________________________

20. Has anyone in the child's family had speech, language or learning problems (for example, the child's mother, father, brothers, sisters or grandparents)?
No _____ Yes _____
↳ If yes, who were they? ________________________________

__The next set of questions is about you.__

21. Your name: ________________________________

22. What is your relationship to the child named on page 1? ____________________

23. In which country were you born?
☐ New Zealand
☐ Australia
☐ England
☐ China (People's Republic of)
☐ India
☐ South Africa
☐ Samoa
☐ Cook Islands
☐ other; please indicate which: ________________________________

__Please continue on the next page.__

24. If you live in New Zealand but were not born here, answer this question:
When did you first arrive to live in New Zealand?
Month (if known) _____ Year _____

25. Which ethnic group do you belong to? Tick the one or ones which apply to you.
☐ New Zealand European
☐ Māori
☐ Samoan
☐ Cook Island Maori
☐ Tongan
☐ Niuean
☐ Chinese
☐ Indian
☐ other such as Dutch, Japanese, Tokelauan; please state: ____________________

26. What is your highest secondary school qualification?
☐ None
☐ NZ School Certificate in one or more subjects *or*
National Certificate level 1 *or*
NCEA level 1
☐ NZ Sixth Form Certificate in one or more subjects *or*
National Certificate level 2 *or*
NZ UE before 1986 in one of more subjects *or*
NCEA level 2
☐ NZ Higher School Certificate *or*
Higher Leaving Certificate *or*
NZ University Bursary / Scholarship *or*
National Certificate Level 3 *or*
NCEA 3 *or*
NZ Scholarship
☐ other secondary school qualification gained in NZ
☐ other secondary school qualification gained overseas

27. Apart from secondary school qualifications, do you have another completed qualification?
No _____ Yes _____
↳ If yes, what is it? ____

28. What is your occupation? ____

29. Today's date is: Day _____ Month _____ Year 20_____

Thank you for your time and effort.

Please return this questionnaire and the *Communicative Development Inventory: Words and Sentences* in the enclosed envelope to:

The Child Language Centre
Department of Communication Disorders
University of Canterbury
Private Bag 4800
Christchurch 8140

APPENDIX C: Parent pack for Part 2

Learning to Talk
Part Two

Information for parents / whānau
An invitation to participate in the next phase of the research project

The research team: Professor Thomas Klee, Professor Stephanie Stokes, Dr Catherine Moran, Jayne Moyle, Brynlea Collin, Louise Hughes and Christine Shalders

Our contact details:
Address: Child Language Centre, 7 Creyke Road, Ilam, Christchurch 8041
Phone: 364 2987 ext. 6431 or 8193. **Cell phone:** 021 0282 7225
Email: ChildLanguageCentre@canterbury.ac.nz
Website: http://www.cmds.canterbury.ac.nz/clc/research/earlyfactors.shtml

Kia ora! Hello!
Thank you for completing the two questionnaires and agreeing to let us contact you again. For the next phase of our research, we would like to see 100 two-year-old children who are not talking much and 100 children who are talking a lot. We would like to assess each child's language, hearing and memory skills now and again in 18 months. We hope you will consider participating in this next phase of the research project.

Before you decide if you want to participate, please take time to read this information. It's important you know why the research is being done and what it will involve. Feel free to ask us if there is anything you are not sure about or would like more information about. Our contact details are above. If you would like to participate further, please call or email us straight away so we can arrange to see you and your child at the Child Language Centre over the next fortnight.

What is the purpose of the project?
Children vary in how quickly they develop speech and language skills. Some start off slowly and then catch up with others their age, while others have persisting difficulties throughout childhood. Currently, we don't know which two year olds will catch up and which ones won't. The purpose of this study is to help us figure out which two year olds we should be concerned about.

Can you help?
We are looking for children who will be aged 24-30 months sometime between now and December 2012. They will need to still be living in Christchurch 18 months from now (although we realise plans can change).

What is involved?
We would like to see your child for an assessment at the Child Language Centre in next fortnight and again 18 months later. We would also like you to fill in another questionnaire about which words your child can say in 3 months' time, so we can see how many more words they have gained over this short period.

What will happen to my child if we take part?

You will need to make four visits to the Child Language Centre (7 Creyke Road). The first two visits will take place in the next fortnight when your child is aged two. The last two visits will take place 18 months after your first visit when your child is aged three. Each visit will take up to 1.5 hours.

When you come to see us, your child will be seen by a speech and language therapist. She will do some short fun tasks with your child to assess their language and memory skills. There are also some activities where your child needs to point to pictures or repeat words. Your child's hearing will be tested with an instrument which is placed in your child's ear briefly. This doesn't hurt and only takes a few minutes. We would also like you to play with your child for 20 minutes. This will all be recorded on video. If your child is unhappy and wants to stop or take a break at any time, this is okay. You will be there with them all the time to see what is happening.

What are the benefits for us?

Your child will receive a free book after the 1st visit as a token of our appreciation. You will receive a $20 voucher after your 2nd and 4th visits for a total of $40. We think your child will enjoy coming to see us, as everything we do is fun and child friendly. You will find it interesting to see where your child is at with their talking. If you have any concerns about your child's speech, you can talk about them with us. We can send you the results from the assessments if you want them. There aren't any risks to you or your child as a result of participating in this study.

What if I change my mind?

You can pull out of the project at any time and you don't have to say why.

What will happen to the results of the project?

We will write up the results and share them with people who work with children across the world. Jayne will also write up her part of the project as a doctoral thesis, which will be available through the University of Canterbury library.

Will my child's name be made public?

No, your child's name won't appear in any publications. Your names and contact details will be noted at the start, but then every child will be given a number, so only the researchers will know which child has which results. The information we have on your child will be kept in secure locked cabinets. Only people working on the project will have access to this information. Videos taken may be used for teaching purposes or at conferences only if you give permission.

Other information:

This project is funded by the Marsden Fund of the Royal Society of New Zealand (research project no. UOC1003, 2011-2014). The project has been reviewed and approved by the University of Canterbury Human Ethics Committee (HEC 2011/121) and the Plunket Ethics Committee.

Consent Form

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Learning to Talk - Part 2

You may cross out any of these statements you do not agree with:

1. I have read and understood the information that was given to me about the research project named above. I have had an opportunity to ask questions and have had them answered. I understand that my participation in this project is voluntary and that we are free to withdraw at any time without giving a reason. I understand that the session will be audio and video recorded. I understand that the information you collect from me will remain confidential and will be stored securely at the university. I understand that any presentations or publications resulting from this project will not refer to us by name. I understand that this project has been reviewed ***and approved*** by the University of Canterbury Human Ethics Committee. On this basis, my child (named below) and I agree to participate in this research project.

2. I agree to also let the researchers use the audio-video recording for **teaching purposes** at the university with the understanding that you will not refer to us by name.

3. I agree to also let the researchers use the audio-video recording at **research conferences** with the understanding that you will not refer to us by name.

MY CHILD'S NAME (please print): ....

My NAME (please print): ....

My signature: ....

Date: ....

❑ I would like you to send me a brief summary of your findings when the study is complete.

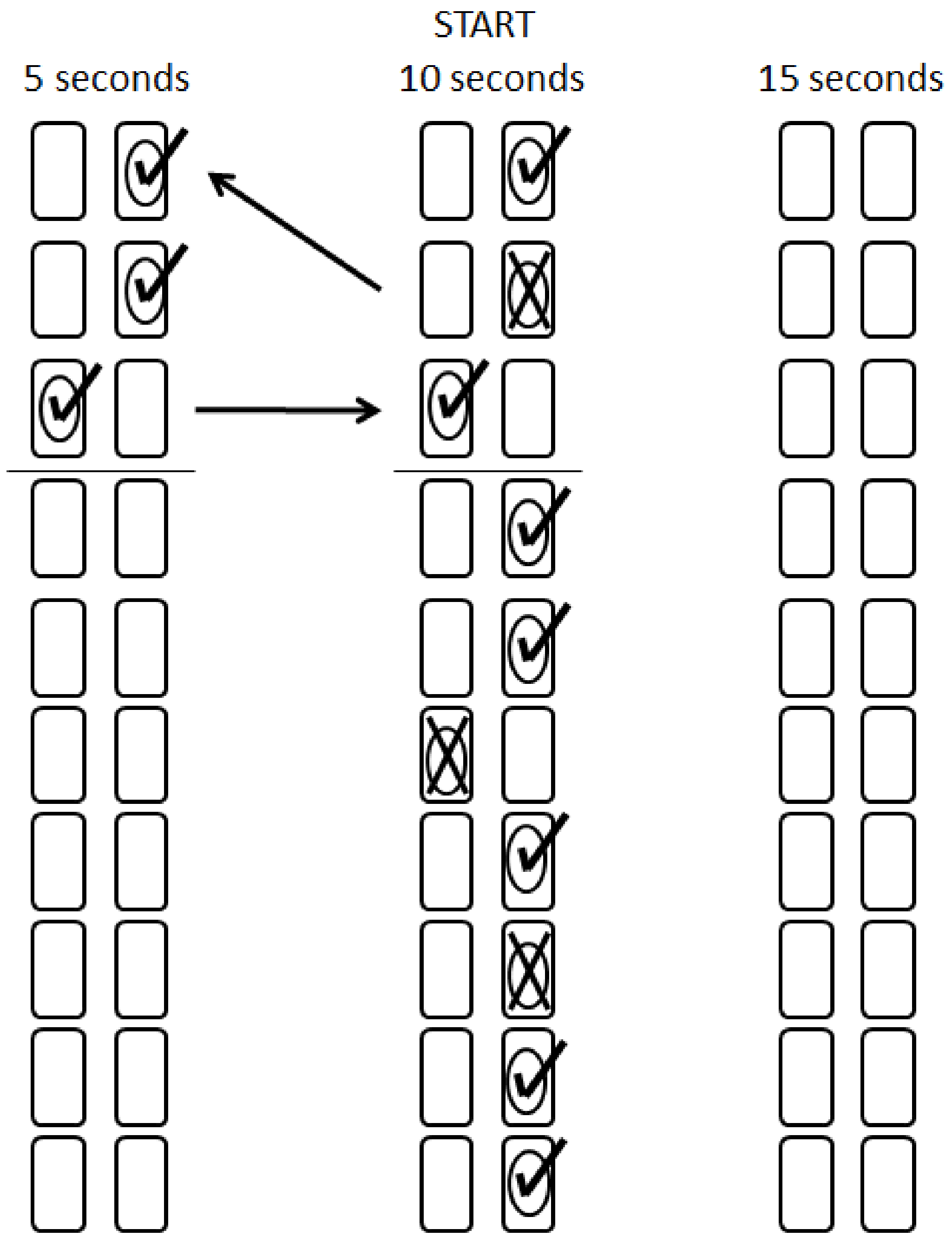
Please contact me by:
❑ Email ____
❑ Cell phone ____
❑ Landline ____
❑ Post

Main Researcher:
Professor Thomas Klee
Email: ChildLanguageCentre@canterbury.ac.nz
Phone: 03 364 2987 ext. 8501

APPENDIX D: Score forms

A not B Task

Participant ID: Joe
Date tested: 16/1/12



Delay at which errors elicited: 10 seconds
Total errors at that level: 3

Total score: 17

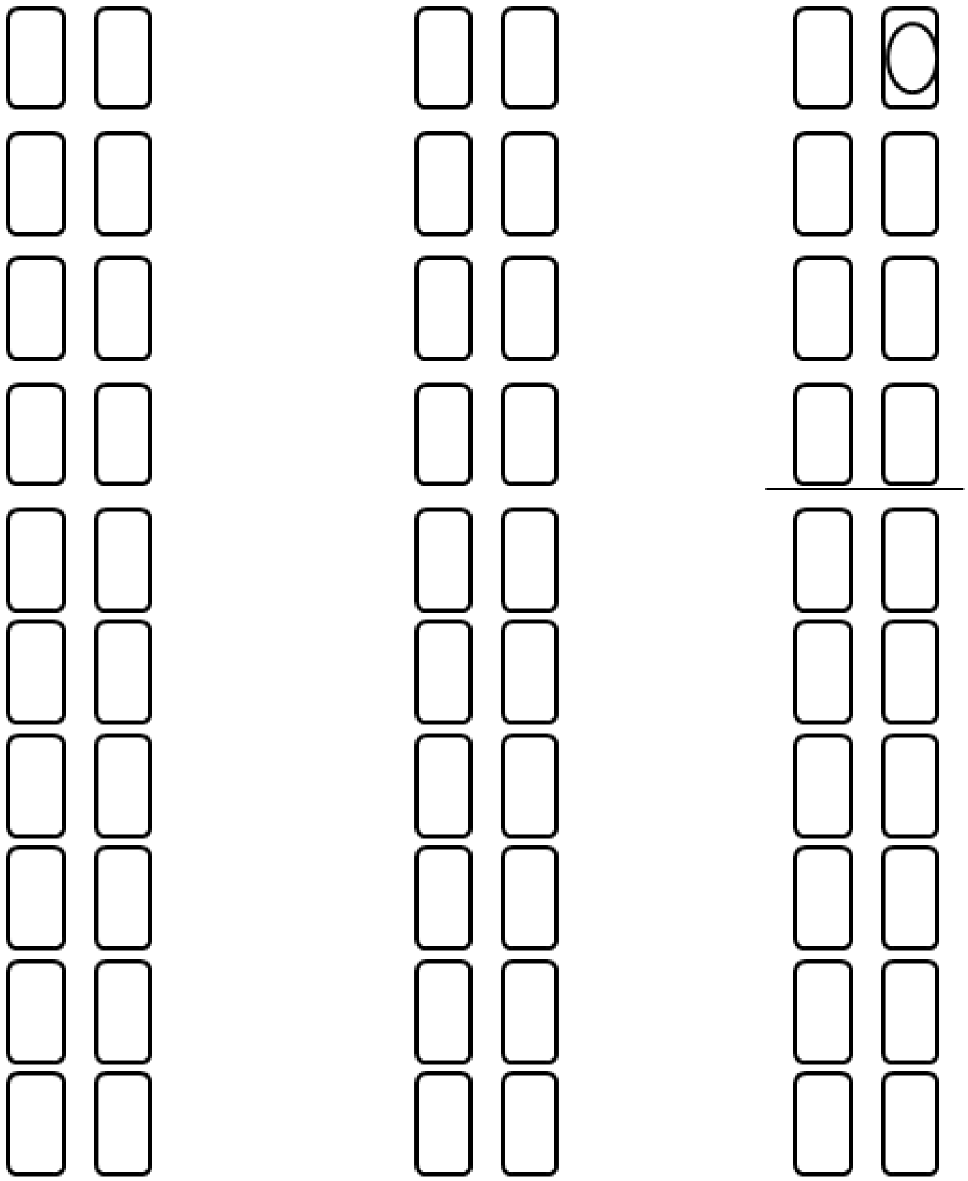
Child must find the reward twice consecutively on the same side before switching.

A not B II Task

Participant ID:
Date tested:

START

5 seconds | 10 seconds | 15 seconds



End here if all correct

Delay at which errors elicited:
Total errors at that level:
Total score:

Child must find the reward thrice consecutively on the same side before switching.

Key Word Working Memory I Task

Participant: Date:

Basal: Two year old children must score three agents correct and three objects correct to proceed. Change items below if child cannot comprehend standard six options.

Ceiling: Discontinue the test if a child misses two key words on every instruction at a key word level, or refuses to comply further.

Allow the child time to play with the toys before beginning to test them. Do not start a trial while the child has any of the toys in their hands. Demonstrate all the actions with several animals before beginning.

Level 1 - one key word / vocabulary testing e.g. Where's the **dog?** Administer until established 3 comprehended agents and objects. This scores a maximum of 3/3, but check they know all the words. As the child to choose from an array of three toys at a time.

cat	**dog**	**pig**	mum	dad	**bed**	**bath**	**drink**	food	book	**Total score**
										/3

Level 2 – two key words e.g. **dog** wants a **drink. Set up array of pig, cat, dog, bed, bath, drink.** Administer 3 trials at this level. The fourth can be administered if a trial was interrupted and discounted.

Instruction	**Describe child's response**	**Item Score**
1. Pig wants the bed		
2. Cat wants a drink		
3. Dog wants the bed		
4. Cat wants the bath		
	TOTAL SCORE:	/6

Level 3 – three key words e.g. pig wants a **drink** and a **bath.**

Instruction	**Describe child's response**	**Item Score**
1. Dog wants a drink and the bed		
2. Pig wants the bath and the food		
3. Cat wants the bed and the bath		
4. Pig wants the food and the bed		
	TOTAL SCORE:	/9

Level 4 – four key words e.g. dog wants a **bath** and **cat** wants the **bed.**

Instruction	**Describe child's response**	**Item Score**
1. Cat wants the bath and pig wants the bed		
2. Dog wants the bed and cat wants the drink		
3. Pig wants the bath and dog wants the drink		
4. Dog wants the bath and cat wants the bed		
	TOTAL SCORE:	/12

Add up subtotals for each level and tally below:

L1	L2	L3	L4	Total Score:
/3	/6	/9	/12	/30

Key word level: The highest item score at any level which the child achieves twice.

Key word level:	

Key Word Working Memory II Task

Participant:
Date:

Basal: Children must score perfect score at a key word level for a basal.

Ceiling: Discontinue testing when child scores 50% or less in each trial at a key word level, or refuses to comply.

Allow the child time to play with the toys before beginning to test them. Do not start a trial while the child has any of the toys in their hands. Demonstrate all the actions with several animals before beginning.

Level 1 - one key word / vocabulary testing e.g. Where's the **dog?** Administer until established 3 comprehended agents and objects. This scores a maximum of 3/3, but check they know all the words. As the child to choose from an array of three toys at a time.

cat	dog	pig	girl	man	bed	bath	drink	food	book	Total score
										/3

Level 2 – two key words e.g. **dog** wants a **drink. Set up array of pig, cat, dog, bed, bath, drink.** Administer 3 trials at this level. The fourth can be administered if a trial was interrupted and discounted.

Instruction	Describe child's response	Item Score
1.Pig wants the bed		
2.Cat wants a drink		
3.Dog wants the bed		
4.Cat wants the bath		
	TOTAL SCORE:	/6

Introduce the food as an addition item to the array of toys.
Level 3 – three key words e.g. pig wants a **drink** and a **bath.**

Instruction	Describe child's response	Item Score
1.Dog wants a drink and the bed		
2.Pig wants the bath and the food		
3.Cat wants the bed and a bath		
4.Pig wants the food and the bed		
	TOTAL SCORE:	/9

Introduce the girl as an additional item to the array of toys.
Level 4 – four key words e.g. dog wants a **bath** and **cat** wants the **bed.**

Instruction	Describe child's response	Item Score
1.Cat wants the bath and girl wants the food		
2.Dog wants a drink and pig wants the bed		
3. Cat wants the food and pig wants a bath		
4. Dog wants the bed and girl wants the drink		
	TOTAL SCORE:	/12

Introduce the book as an additional item to the array of toys.
Level 5: five key words e.g. cat wants the **food , dog** wants the **bath** and the **bed**

Instruction	Describe child's response	Item Score
1. Pig wants the bath and girl wants the bed and the food		
2. Dog wants the book and cat wants the bath and the drink		
3. Girl wants the bath and pig wants the drink and the book		
4. Cat wants the food and girl wants the bed and the bath		
	TOTAL SCORE:	/15

Introduce the man as an additional item to the array of toys.
Level 6: Six key words e.g. pig wants the **food, girl** wants the **book and dog** wants a **bath.**

Instruction	Describe child's response	Item Score
1. Dog wants the bath, man wants the food and pig wants the bed		
2. Cat wants the book, dog wants the bath and girl wants the drink		
3. Girl wants the food, pig wants the bed and man wants the bath		
4. Dog wants the drink, cat wants the bed and pig wants the book.		
	TOTAL SCORE:	/18

Add up subtotals for each level and tally below:

L1	L2	L3	L4	L5	L6	Total Score:
/3	/6	/9	/12	/15	/18	/63

Key word level: The highest item score at any level which the child achieves twice.

Key word level:	

Test of Early Non-Word Repetition (TENR)

Participant number: ________________________ Date: ___________

IPA target	IPA child response	Score	Target score
mad			3
neit			3
paim			3
bouk			3
kou gə			4
da fi			4
lɜ pou			4
fu pɪm			5
mou kɜ ri			6
dou pɜ lut			7
bæ lɜ kɒn			7
fi sai mɒt			7
pɜ du lə meip			9
fɛ nɜ rai sɛk			9
wu gɜ læ mɪk			9
lɒ dɜ næ tɪʃ			9
gi lɜ ma fu kou			10
lɜ tei di ku nei			10
gɔ lu mɜ fi nai			10
ba fu mou wu di			10
	TOTAL SCORE		132

APPENDIX E: Looking While Listening I task coding procedures

Elan Template

Once the child had completed the task, the video and sound files were opened in Elan (Max Planck Institute for Psycholinguistics, 2012). An Elan template was made for analysis of the data which consisted of six tiers:

- "Usable" – every recorded trial was marked "usable" or "unusable". This was determined by whether the child was looking at the screen in the first 40ms of the Word Window. This was a parent tier, so each tier below this one was exactly the same length and started at the same time in the video.
- "Word window" – this was the two second period of time from the start of the target word
- "Word Looking" – every 40ms period in the word window was marked by where the child was looking. The four categories were "left", "right", "away" or "shift".
- "Initial looking" – each word window was marked whether the child was looking at the target or distractor picture in the first 40ms of the word window.
- "Latency" – each trial which was marked "distractor" in the Initial Looking tier, was coded for latency. This was the period of time from the start of the word window until the first frame when the child shifted their gaze to the target picture.
- "Comments" – an opportunity to note special circumstances or coding decisions

Determining the start of the Word Window:

Elan software does not allow time markings on their templates, meaning it was not possible to make a master template for the start of each target word. Initially (for the first 45 participants), the onset of the target word was determined perceptually using the Elan auditory play-back feature and intensity information displayed visually. However, while recoding some trials to check for reliability, I noted some inconsistencies in the marking of the onset of the word window (10-50ms difference between codings). In some cases, this changed the coding of the eye movements by a frame, which in turn had a minor impact on the mean latency. These problems arose because of the difficulty perceiving the precise onset of each word using intensity and auditory perception alone (especially for plosive-initial words). I decided to use visual analysis of the waveform of the audio files from the original recordings instead. The start of “where’s” was comparatively easy to determine using the intensity display on Elan. The same time delay from the start of the “where’s” to the start of each target word could be used for every occurrence of that audio file in the slide show, thus also increasing consistency of coding.

Therefore the onset of target words on the original audio files was determined by visual and acoustic analysis of the waveform using PRAAT software (Boersma, 2001). All audio files were marked with the time from start of the “where’s” to start of the target word by myself. Target words either started with a vowel, plosive or fricative. For the vowel, the end of “the” and just before the glottal before “elephant” was marked as the word onset. For fricatives, the end of the second formant (for unvoiced plosives) and beginning of friction pattern was marked. For plosives, the end of the second formant and the beginning of the up down pattern in intensity at the start of the closure was marked. When there were two up down patterns, the first was taken as the start of closure. This represents the start of the closure of the plosive, not the onset of the burst. Thus silent stop information was included as part of the start of the target words for plosive initial words. All onsets were checked by an expert in spectral analysis (Dr Margaret MacLaghan) and were found to be accurate to within 20ms. These onset times were then used for coding of word windows for the remaining participants (P46 - P83). Target word onset codings for P1- P45 were checked against the new onset times and were changed if the difference was 40ms or greater from the original coding. This figure was decided upon as any smaller difference in the word onset does not affect visual information recorded digitally at 40ms intervals.

Coding procedures for each trial

The list describes how each trial is to be coded using the template on Elan. Firstly the action to be taken on each tier is described, then the criteria for decision making around the codings. The rationale for the action is also briefly described for each tier.

Tier1 : Usable

Action: Mark all trials as either "usable" or "unusable".

Criteria for "unusable":

- The child is not looking at the screen at the onset of the target word (away).
- Extra noise in the audio recording means the start of the "where's" cannot be located precisely.
- The child is learning the task i.e. all first first trials are coded as "unusable". The next few trials may also be marked as "unusable" for some children, if the child is not responding reliably to the task yet. "Learning the task" is defined as *either* shifting correctly from distractor to target for a distractor initial trial; *or* staying on the target for target initial trial *as well as* looking at the target picture for longer than the distractor over the two second word window; *or* the child may show by body language or verbalisations that they understand the task.
- The child is trying to leave the task or is inattentive as shown by off-task behaviour e.g. making noise / squirming / kicking the desk.
- If the child's eyes are covered at a critical point meaning looking direction cannot be coded.
- Sometimes children appear to "zone out" while still looking at the screen – if a child

is not looking directly at either the left or right picture and does not change their eye gaze throughout the Word Window, this may also coded as “unusable”.

Rationale: No further time should be spent on trials which cannot be used for the latency measure. Counts for trial number and number of usable trials can be gained from this tier.

Tier 2: Word Window

Action: Mark the two seconds after onset of target word for all usable trials

Criteria: Measure from the onset of “where’s” (determined visually using the intensity graph) to the onset of the target word (length of interval determined using spectral analysis of the word form referred to earlier).

Rationale: This sets the two second period in which the looking direction will be annotated.

Tier 3: Word Looking

Controlled vocabulary: left, right, shift, away

Action: Annotate each 40ms frame within the two second word window with the child’s looking direction.

Criteria:

- **Left** – child is looking at left picture (“left” from the video viewer’s perspective, not the child’s). The child’s eyes may shift around within the picture.
- **Right** – as above, but child is looking at the right picture.
- **Shift** – the child’s eyes are shifting between the two pictures. Start coding from first frame where the eyes move. Shifts are usually 1-3 frames in duration. Often just after a shift, the child’s head still moves for one or two frames but the eyes are fixed on the picture: Code these frames as “right” / “left”, not as “shift”.
- **Away** – the child looking anywhere but the two pictures or between them.

- **Blinking:** Consider where the child was looking immediately prior to and after the blink. If their eyes remain in the same position (e.g. left – blink – left), code the blink time as “left”. If the child shifted in a blink (e.g. right – blink – left), then code the blink as “shift”. If the child is looking elsewhere before or after the blink (e.g. away – blink – left), then code the frames where the eyes are closed as “away”.

Rationale: This information is critical for measuring the latency of shift between distractor and target pictures

Tier 4: Initial Looking

Action: Determine which trials are target- or distractor-initial.

Controlled vocabulary: Target, distractor, shift, away.

Criteria: Check the looking direction for the first frame. Compare with the reference sheet to see if the target word was on the left or right for this trial. If the picture the child was looking at is the target, mark as “target-initial”, if it was the distractor, mark as “distractor-initial”. If the first frame for the Word Looking tier was marked “shift” or “away”, mark this tier in the same way.

Rationale: To be able to identify “distractor-initial” trials easily to calculate mean latencies.

Tier 5: Latency

Action: Mark the latency from the start of the word window until the initiation of the shift to the target picture for distractor initial trials.

Criteria: If this trial was marked as “distractor-initial” and the child shifted to the target picture within 240-1000ms after the start of the word window. If not, leave this tier blank.

Rationale: To measure latency of shift from distractor to target.

Tier 6: Comments

Controlled vocabulary: None

Action: Note rationale for coding decisions for later reference

Criteria: Use this tier when there is possible ambiguity between coders in decisions made. This will mainly affect whether trials are classed as “usable” or “unusable” and how blinks are coded.

Rationale: To be able to track how decisions were made to reduce inconsistencies.

APPENDIX F: Diagnostics for the three multiple linear regression models

There are three multiple linear regression models reported in this thesis. The same diagnostic procedure was carried out on each one to check the underlying assumptions of the models. The three key assumptions to check are: 1) linearity (at every possible value of the dependent variable the expected value of the residuals is zero) 2) homoscedasticity (the variance of the residuals is the same at every set of values of the independent variable) and 3) normality of residuals (at each value of the dependent variable, the distribution of residuals is normal) (Miles & Shevlin, 2001).

1. Multiple linear regression model predicting CDI scores at 24-30 months using concurrent predictors (Section 4.3.3)

There may have been some collinearity impacting on the model, as the TENR and PCC correlated $r(79) = .78$ ($p < .001$). For this reason, the model was also calculated without PCC. This analysis demonstrated that there was a large amount of shared variance between these two measures and the dependent variable as stated in Section 4.3.3. This overlapping variance was considered when interpreting this regression model. The scatterplot of the standardised residuals and predicted values looked like a random array of dots spread out, therefore the assumptions of linearity and homoscedasticity were met. The assumption of normality of residuals was met as the P-P plot of standardised residuals looked like a nearly straight line on a 45 degree angle.

The data were inspected to ensure no case departed significantly from the model or had an undue influence. Firstly the amount of error in the model for individual cases was checked. One case showed a standardised residual of just over 1.96 (Participant 72). Five percent of cases are expected to fall in this range according to a normal distribution so this is lower than expected (four cases for 77 participants). No cases had a standardised residual of over 2.2. Secondly, the influence which individual cases had on the model was checked. Cook's distances were all under one as recommended (Miles & Shevlin, 2001). Expected leverage is (k+1) / n, where k is the number of predictors and n is the sample size. In this case expected leverage is (8+1) / 77 = 0.12. Cases having a leverage of over two times this value should be inspected (Miles & Shevlin, 2001). A high leverage combined with a high residual can be a problem, having both a high degree of error and also a high influence on the model (Miles & Shevlin, 2001). There were three cases with leverage values over 0.23 (Participants 30, 54, 82). As none of these cases also had a high residual, there was no need to consider removing them from the model.

This diagnostic procedure was followed for all three multiple regression models in the current research. From this point on, the diagnostics are not described in detail for each regression model, but instead are summarised below.

2. Multivariate regression model for expressive language at 41-49 months using concurrent predictors (Section 6.3.3)

There should be no difficulties with collinearity as none of the predictor variables in the final model correlated with each other highly. The standardised residuals basically followed the 45 degree angle in the P-P plot, but ran slightly above it, meaning the predicted values ran above what was observed. Therefore bootstrapping was performed to check the p values of the coefficients. Bootstrapping resamples with replacement all the observations in the data and fits a regression line to each one. From this series of analyses, a distribution of the beta coefficients is calculated. This is used to estimate the standard error of the coefficients, which is then used to create confidence intervals and significance tests. All the p values for the coefficients remained less than the alpha level set (.05), except for age which was only left in the model as a control. Therefore we can be confident that the beta values differ from zero in this model, with the likely exception of age. Only the regression model without bootstrapping was reported in this thesis, as there were no major differences between the two models.

Cook's distances were all under one. There was one residual over 1.96, fewer than the amount expected for a sample size of 69. None of the residuals were over three. In this case expected leverage was (6+1) / 69 = 0.10. There were two cases with a leverage value of over 0.20 (Participants 3, 79). Neither case had a high residual.

3. Multivariate regression model for total language scores at 41-49 months using early predictors (Section 7.2.2)

There is unlikely to be a problem with multicollinearity. Initial KWM and PLS-4 AC raw scores correlated with each other by $r = .73$ ($p < .001$), however they both predicted unique variance in the model, and the standard errors of beta were small. Emotional Control (log) and Shift (log) correlated on a bivariate level at $r = .69$ (p < .001). The standard errors of these coefficients were higher relative to the beta value meaning, there is less confidence that these values gained from the sample represent the population well. However the Variance Inflation Factor (VIF) values were all below 2.5 and the tolerance statistics were all higher than .4, meaning there is no need for concern about multicollinearity (Field, 2009).

The standardised residuals in the P-P plot are basically following the 45 degree angle, but ran above it for less than half of the observed values. Therefore bootstrapping was performed as a check on the accuracy of the *p* values for the coefficients in the model. All the *p* values for the coefficients remained significant at $p < .05$. Therefore we can have confidence that the beta values are significantly different from zero in this model. Only the regression model without bootstrapping was reported in this thesis, as there were no major differences between the two models.

Cook's distances ranged from 0.00 to 0.10. In this case expected leverage was (7+1) / 76 = 0.11. There were three cases with leverage values over 0.22 (P10, 82, 60). There were three standardised residuals over 1.96 (P5, P11 and P41); which is within what is expected for a sample size of 76. None of these cases had both a high leverage value and a high residual.

APPENDIX G:

Parent Questionnaire II

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Please complete this questionnaire and hand it in before you leave today. You may also complete it at home and return it at your next visit. If you'd rather talk to us about the questions instead of filling out this form, just let us know!

1. Child's name: ____

2. Your name: ____

3. Your relationship to the child: ____________________________________

4. If any other adults live in the child's home apart from you, what is their relationship to the child?

   ____

5. Child's birth date: Day _____ Month _____ Year 20_____

6. Child's birth weight: __________ grams (*or* _____lbs _____oz)

7. Which hand does the child prefer to use when holding a spoon or crayon?

   Left ___ Right ___ Either ___

8. Was the child born prematurely? No _____ Yes _____
   ↳ If yes, by how many weeks? _______

9. Has the child had any health problems, including ear infections, during the past 18 months?

   No _____ Yes _____
   ↳ If yes, what are they? ____________________________________

   ____

   ____

10. Number of children in family, *including* this child: _____

11. Are any other languages spoken in the child's home?

    No _____ Yes _____
    ↳ If yes, which ones? ____________________________________
    ↳ What percentage of the time? ______________________________

12. Is the child in day care or an early childhood education programme or cared for regularly by anyone else?

    No _____ Yes _____
    ↳ If yes, how many hours per week on average? _____

13. Did the child receive any speech and language intervention during the last 18 months?

No _____ Yes _____
↳ If yes, please describe. ______________________________________

14. Has anyone in the child's family had speech, language or learning problems (for example, the child's mother, father, brothers, sisters or grandparents)?
No _____ Yes _____
↳ If yes, who were they? ______________________________________

15. Do you have any concerns about the child's ability to hear? Yes ___ No ___

16. Do you have any concerns about the child's language development? Yes ___ No ___

17. Do you have any concerns about the child's ability to communicate? Yes ___ No ___

18. If you answered "Yes" to any of the last 3 questions, say why you are concerned:

____

____

____

____

19. What is your occupation? ______________________________________

20. Today's date is: Day _____ Month _____ Year 20_____

21. Your home phone number: ______________________________________

22. Your cell phone number: ______________________________________

23. Your email address: ______________________________________

24. Your postal address:

____

____

____

____

Thank you very much for answering this.

The Child Language Centre
Department of Communication Disorders
University of Canterbury
Private Bag 4800
Christchurch 8140
Phone: 364 2987, ext 8317 or 8501
Email: ChildLanguageCentre@canterbury.ac.nz

References

Adams, A., & Gathercole, S. E. (1995). Phonological working memory and speech production in preschool children. *Journal of Speech & Hearing Research, 38*(2), 403-414.

Alcock, K. J., & Krawczyk, K. (2010). Individual differences in language development: Relationship with motor skill at 21 months. *Developmental Science, 13*, 677-691. doi: 10.1111/j.1467-7687.2009.00924.x

Alloway, T. P. (2012). *Automated Working Memory Assessment-II (AWMA-II)*. London: Pearson Assessment.

Alloway, T. P., & Gathercole, S. (Eds.). (2006). *Working memory and neurodevelopmental disorders*. New York, NY: Psychology Press.

Archibald, L., & Gathercole, S. (2007). Non-word repetition in specific language impairment: More than a phonological short-term memory deficit. *Psychonomic Bulletin & Review, 14*, 919.

Archibald, L., & Gathercole, S. E. (2006a). Short-term and working memory in specific language impairment. *International Journal of Language & Communication Disorders, 41*, 675-693. doi: 10.1080/13682820500442602

Archibald, L., & Gathercole, S. E. (2006b). Visuospatial immediate memory in specific language impairment. *Journal of Speech, Language & Hearing Research, 49*, 265-277. doi: 10.1044/1092-4388(2006/022)

Aslin, R. N., & Newport, E. L. (2009). What statistical learning can and can't tell us about language acquisition. In J. Colombo, P. McCardle & L. Freund (Eds.), *Infant pathways to language: Methods, models, and research disorders.* (pp. 15-29). New York, NY US: Psychology Press.

Baddeley, A. (2007). *Working memory, thought and action*. New York, NY: Oxford University Press.

Baddeley, A. (2012). Working Memory: Theories, Models, and Controversies. *Annual Review of Psychology, 63*(1), 1-29. doi: 10.1146/annurev-psych-120710-100422

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105*, 158-173. doi: 10.1037/0033-295x.105.1.158

Baddeley, A., Papagno, C., & Vallar, G. (1988). When long-term learning depends on short-term storage. *Journal of Memory and Language, 27*, 586-595. doi: 10.1016/0749-596X(88)90028-9

Bailey, T. M., & Plunkett, K. (2002). Phonological specificity in early words. *Cognitive Development, 17*, 1265-1282. doi: 10.1016/S0885-2014(02)00116-8

Baird, G., Dworzynski, K., Slonims, V., & Simonoff, E. (2010). Memory impairment in children with language impairment. *Developmental Medicine & Child Neurology, 52*, 535-540. doi: 10.1111/j.1469-8749.2009.03494.x

Ballem, K. D., & Plunkett, K. (2005). Phonological specificity in children at 1;2. *Journal of Child Language, 32*, 159-173.

Bayley, N. (1969). *Scales of infant mental development*. New York, NY: Psychological Corp.

Bayliss, D., Jarrold, C., Baddeley, A., Gunn, D., & Leigh, E. (2005). Mapping the developmental constraints on working memory span performance. *Developmental Psychology, 41*, 579-597. doi: 10.1037/0012-1649.41.4.579

Beckage, N., Smith, L., & Hills, T. (2011). Small worlds and semantic network growth in typical and late talkers. *PLoS ONE, 6*(5), e19348. doi: 10.1371/journal.pone.0019348

Beitchman, J. H., Nair, R., Clegg, M., Ferguson, B., & Patel, P. G. (1986). Prevalence of psychiatric disorders in children with speech and language disorders. *Journal of the American Academy of Child Psychiatry, 25*, 528-535.

Beitchman, J. H., Nair, R., Clegg, M., & Patel, P. G. (1986). Prevalence of speech and language disorders in 5-year-old kindergarten children in the Ottawa-Carleton region. *Journal of Speech and Hearing Disorders, 51*, 98-110.

Bishop, D. V. M. (2006). Developmental cognitive genetics: How psychology can inform genetics and vice versa. *The Quarterly Journal of Experimental Psychology, 59*, 1153 - 1168.

Bishop, D. V. M., Adams, C. V. & Norbury, C. F. (2006). Distinct genetic influences on grammar and phonological short term memory deficits: Evidence from 6-year-old twins. *Genes, Brain and Behaviour, 5*, 158-169.

Bishop, D. V. M., Bishop, S., Bright, P., James, C., Delaney, T., & Tallal, P. (1999). Different origin of auditory and phonological processing problems in children with language impairment: evidence from a twin study. *Journal of Speech Language Hearing Research, 42*, 155-168.

Bishop, D. V. M., Carlyon, R., Deeks, J., & Bishop, S. (1999). Auditory temporal processing impairment: neither necessary nor sufficient for causing language impairment in children. *Journal of Speech Language Hearing Research, 42*, 1295-1310.

Bishop, D. V. M., Holt, G., Line, E., McDonald, D., McDonald, S.,& Watt, H. (2012). Parental phonological memory contributes to prediction of outcome of late talkers from 20 months to four years: A longitudinal study of precursors of specific language imairment. *Journal of Neurodevelopmental Disorders, 4*, 1-12.

Bishop, D. V. M., & Norbury, C. F. (2005a). Executive functions in children with communication impairments, in relation to autistic symptomatology: 2: Response inhibition. *Autism, 9*, 29-43. doi: 10.1177/1362361305049028

Bishop, D. V. M., & Norbury, C. F. (2005b). Executive functions in children with communication impairments, in relation to autistic symptomatology: I: Generativity. *Autism, 9*, 7-27. doi: 10.1177/1362361305049027

Bishop, D. V. M., North, T., & Donlan, C. (1996). Nonword repetition as a behavioural marker for inherited language impairment: Evidence from a twin study. *Journal of Child Psychology and Psychiatry, 37*, 391-403. doi: 10.1111/j.1469-7610.1996.tb01420.x

Blake, J., Austin, W., Cannon, M., Lisus, A., & Vaughan, A. (1994). The relationship between memory span and measures of imitative and spontaneous language complexity in preschool children. *International Journal of Behavioral Development, 17*, 91-107. doi: 10.1177/016502549401700106

Boersma, P. (2001). Praat: A system for doing phonetics by computer. *Glot International, 5*, 341-345.

Bonifacio, S., Girolametto, L., Bulligan, M., Callegari, M., Vignola, S., & Zocconi, E. (2007). Assertive and responsive conversational skills of Italian-speaking late talkers. *International Journal of Language & Communication Disorders, 42*, 607-623. doi: 10.1080/13682820601084386

Botting, N., & Conti-Ramsden, G. (2001). Non-word repetition and language development in children with specific language impairment (SLI). *International Journal of Language & Communication Disorders, 36*, 421-432. doi: 10.1080/13682820110074971

Bowey, J. A. (2001). Non-word repetition and young children's receptive vocabulary: A longitudinal study. *Applied Psycholinguistics, 22*, 441-469.

Boyle, W., Lindell, A. K., & Kidd, E. (2013). Investigating the role of verbal working memory in young children's sentence comprehension. *Language Learning, 63*, 211-242. doi: 10.1111/lang.12003

Bricker, D., & Squires, J. (1999). *Ages and Stages Questionnaires: a parent completed child monitoring system.* (2nd ed.). Baltimore: Paul H. Brookes.

Briscoe, J., & Rankin, P. M. (2009). Exploration of a 'double-jeopardy' hypothesis within working memory profiles for children with specific language impairment. *International Journal of Language & Communication Disorders, 44*, 236-250. doi: 10.1080/13682820802028760

Burgess, P. W., Alderman, N., Forbes, C., Costello, A., M-A.Coates, L., Dawson, D. R., . . . Channon, S. (2006). The case for the development and use of measures of executive function in experimental and clinical neuropsychology. *Journal of the International Neuropsychological Society, 12*, 194-209. doi: 10.1017/S1355617706060310

Buschmann, A., Jooss, B., Rupp, A., Dockter, S., Blaschtikowitz, H., Heggen, I., & Pietz, J. (2008). Children with developmental language delay at 24 months of age: results of a diagnostic work-up. *Developmental Medicine & Child Neurology, 50*, 223-229. doi: 10.1111/j.1469-8749.2008.02034.x

Caplan, D., & Waters, G. (2013). Memory mechanisms supporting syntactic comprehension. *Psychonomic Bulletin & Review, 20*, 243-268. doi: 10.3758/s13423-012-0369-9

Catts, H. W., Adolf, S. M., Hogan, T. P., & Ellis Weismer, S. (2005). Are specific language impairment and dyslexia distinct disorders? *Journal of Speech, Language & Hearing Research, 48*, 1378-1396.

Chiat, S. (2001). Mapping theories of developmental language impairment: Premises, predictions and evidence. *Language & Cognitive Processes, 16*, 113-142. doi: 10.1080/01690960042000012

Chiat, S., & Roy, P. (2007). The Preschool Repetition Test: An evaluation of performance in typically developing and clinically referred children. *Journal of Speech, Language, and Hearing Research, 50*, 429-443.

Chiat, S., & Roy, P. (2008). Early phonological and sociocognitive skills as predictors of later language and social communication outcomes. *Journal of Child Psychology and Psychiatry, 49*, 635-645. doi: 10.1111/j.1469-7610.2008.01881.x

Chiat, S., & Roy, P. (2013). Early predictors of language and social communication impairments at ages 9-11 years: A follow-up study of early-referred children. *Journal of Speech, Language & Hearing Research, 56*, 1824-1836. doi: 10.1044/1092-4388(2013/12-0249)

Choudhury, N., & Benasich, A. A. (2003). A family aggregation study: the influence of family history and other risk factors on language development. *Journal of Speech, Language & Hearing Research, 46*, 261-272.

Cleary, J. E. (2002). *The predictive value of young children's prelinguistic communication and motor skills.* Doctoral dissertation. Retrieved from ProQuest LLC. (3043345)

Clegg, J., Hollis, C., Mawhood, L., & Rutter, M. (2005). Developmental language disorders - A follow-up in later adult life. Cognitive, language and psychosocial outcomes. *Journal of Child Psychology and Psychiatry, 46*, 128-149. doi: 10.1111/j.1469-7610.2004.00342.x

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale: Lawrence Erlbaum.

Conboy, B. T., Sommerville, J. A., & Kuhl, P. K. (2008). Cognitive control factors in speech perception at 11 months. *Developmental Psychology, 44*, 1505-1512. doi: 10.1037/a0012975

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? *Progress in Brain Research, 169,* 323-338. doi: 10.1016/S0079-6123(07)00020-9

Cowan, N., Rouder, J. N., Blume, C. L., & Saults, J. S. (2012). Models of verbal working memory capacity: What does it take to make them work? *Psychological Review, 119*, 480-499. doi: 10.1037/a0027791

Creusere, M., Alt, M., & Plante, E. (2004). Recognition of vocal and facial cues to affect in language-impaired and normally-developing preschoolers. *Journal of Communication Disorders, 37,* 5-20. doi: 10.1016/s0021-9924(03)00036-4

D'Odorico, L., Assanelli, A., Franco, F., & Jacob, V. (2007). A follow-up study on Italian late talkers: Development of language, short-term memory, phonological awareness, impulsiveness, and attention. *Applied Psycholinguistics, 28*, 157-169. doi: 10.1017/s0142716407070087

Dale, P. S., & Fenson, L. (1996). Lexical development norms for young children. *Behavioral Research Methods, Instruments, & Computers, 28*, 125-127.

Dale, P. S., Price, T., Bishop, D. V. M., & Plomin, R. (2003). Outcomes of early language delay: I. Predicting persistent and transient language difficulties at 3 and 4 years. *Journal of Speech, Language & Hearing Research, 46*, 544-560.

Dale, P. S., Simonoff, E., Bishop, D. V. M., Eley, T. C., Oliver, B., Price, T., . . . Plomin, R. (1998). Genetic influence on language delay in two-year-old children. *Natural Neuroscience, 1,* 324-328. doi: 10.1038/1142

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19,* 450-466. doi: 10.1016/s0022-5371(80)90312-6 de Koning, H. J., de Ridder-Sluiter, J. G., van Agt, H. M. E., & Reep-van den Bergh, C. M. M. (2004). A cluster-randomised trial of screening for language disorders in toddlers. *Journal of Medical Screening, 11*, 109-116.

Desmarais, C., Sylvestre, A., Meyer, F., Bairati, I., & Rouleau, N. (2008). Systematic review of the literature on characteristics of late-talking toddlers. *International Journal of Language & Communication Disorders, 43*, 361-389. doi: 10.1080/13682820701546854

Diamond, A. (1990). The development and neural bases of memory functions as indexed by the AB and delayed response tasks in human infants and infant monkeys. *Annals of the New York Academy of Sciences*, *608*, 267-317. doi: 10.1111/j.1749-6632

Dibbets, P., Bakker, K., & Jolles, J. (2006). Functional MRI of task switching in children with specific language impairment (SLI). *Neurocase, 12*, 71-79.

Dodd, B., & Crosbie, S. (2002). Language and cognition: Evidence from disordered language. In U. Goswani (Ed.), Blackwell Handbook of Childhood Cognitive Development (pp. 470-490): Blackwell Publishers Ltd. doi: 10.1002/9780470996652.ch21

Dodd, B., Hua, Z., Crosbie, S., Holm, A., & Ozanne, A. (2002). *Diagnostic Evaluation of Articulation and Phonology (DEAP)*. London, UK: Pearson Assessment.

Dodd, B., & McIntosh, B. (2010). Two-year-old phonology: Impact of input, motor and cognitive abilities on development. *Journal of Child Language, 37*, 1027-1046. doi: 10.1017/S0305000909990171

Dollaghan, C. (2013). Late talker as a clinical category: A critical evaluation. In L. Rescorla & P. S. Dale (Eds.), *Late talkers: Language Development, Interventions and Outcomes* (pp. 91-112). Baltimore, Maryland: Paul H. Brookes Publishing Co.

Dollaghan, C., & Campbell, T. E. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research, 41*, 1136-1146.

Durkin, K., & Conti-Ramsden, G. (2010). Young people with specific language impairment: A review of social and emotional functioning in adolescence. *Child Language Teaching and Therapy, 26*, 105-121. doi: 10.1177/0265659010368750

Ellis, E. M., & Thal, D. J. (2008). Early language delay and risk for language impairment. *Perspectives on Language Learning and Education, 15*, 93-100. doi: 10.1044/lle15.3.93

Ellis Weismer, S. (2007). Typical talkers, late talkers and children with specific language impairment: A language endowment spectrum? In R. Paul (Ed.), *Language disorders from a developmental perspective. Essays in honour of Robin S. Chapman*. NJ: Lawrence Erlbaum.

Ellis Weismer, S., & Evans, J. (2002). The role of processing limitations in early identification of specific language impairment. *Topics in Language Disorders, 22*, 15-29. doi: 10.1097/00011363-200205000-00004

Ellis Weismer, S., Evans, J., & Hesketh, L. J. (1999). An examination of verbal working memory capacity in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 42*, 1249-1260.

Ellis Weismer, S., & Hesketh, L. (1996). Lexical learning by children with specific language impairment: Effects of linguistic input presented at varying speaking rates. *Journal of Speech and Hearing Research, 39*, 177-190.

Ellis Weismer, S., Murray-Branch, J., & Miller, J. F. (1994). A prospective longitudinal study of language development in late talkers. *Journal of Speech and Hearing Research, 37*, 852-867.

Epsy, K. A., Bull, R., Kaiser, H., Martin, J., & Banet, M. (2008). Methodological and conceptual issues in understanding the development of executive control in the preschool period. In V. Anderson, R. Jacobs & P. J. Anderson (Eds.), *Executive functions and the frontal lobes; a lifespan perspective; studies in neuropsychology, neurology and cognition* (pp. 105-121). New York: Taylor & Francis.

Epsy, K. A., Kaufmann, P., McDiarmid, M., & Glisky, M. (1999). Executive functioning in preschool children: Performance on A Not B and other delayed response format tasks. *Brain and Cognition*, *41*, 178-199. doi: 10.1006/brcg.1999.1117

Epsy, K. A., Kaufmann, P. M., Glisky, M. L., & McDiarmid, M. D. (2001). New procedures to assess executive functions in preschool children. *Clinical Neuropsychologist*, *15*, 46-58.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102*, 211-245. doi: 10.1037/0033-295x.102.2.211

Ewing-Cobbs, L., Prasad, M. R., Landry, S. H., Kramer, L., & DeLeon, R. (2004). Executive functions following traumatic brain injury in young children: A preliminary analysis. *Developmental Neuropsychology*, *26*, 487-512.

Fenson, L., Marchman, V., Thal, D., Dale, P. S., Reznick, J. S., & Bates, E. (2007). *MacArthur-Bates Communicative Development Inventories* (2nd ed.). Baltimore, Md: Paul H. Brookes Pub. Co.

Fernald, A., & Marchman, V. A. (2012). Individual differences in lexical processing at 18 months predict vocabulary growth in typically developing and late-talking toddlers. *Child Development, 83,* 203-222. doi: 10.1111/j.1467-8624.2011.01692.x

Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology, 42,* 98-116. doi: 10.1037/0012-1649.42.1.98

Fernald, A., Pinto, J. P., Swingley, D., Weinberg, A., & McRoberts, G. W. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science, 9*, 228-231. doi: 10.1111/1467-9280.00044

Fernald, A., Zangl, R., Portillo, A. L., & Marchman, V. (2008). Using eye movements to monitor spoken language comprehension by infants and young children. In I. A. Sekerina, E. M. Fernandez & H. Clahsen (Eds.), *Developmental Psycholinguistics: Online methods in children's language processing*. Amsterdam: John Benjamins Publishing Company.

Field, A. (2009). *Discovering Statistics Using SPSS* (3rd ed.). London: Sage Publications Inc.

Finneran, D., Francis, A., & Leonard, L. (2009). Sustained attention in children with specific language impairment (SLI). *Journal of Speech, Language and Hearing Research, 52*, 915-929. doi: 1092-4388_2009_07-0053

Fisher, S. E. (2007). Molecular windows into speech and language disorders. *Folia Phoniatrica et Logopaedica, 59,* 130-140.

Fisher, S. E., Vargha-Khadem, F., Watkins, K. E., Monaco, A. P., & Pembrey, M. E. (1998). Localisation of a gene implicated in a severe speech and language disorder. *Nature Genetics, 18,* 168-170. doi: 10.1038/ng0298-168

Florit, E., Roch, M., & Levorato, M. C. (2011). Listening text comprehension of explicit and implicit information in preschoolers: The role of verbal and inferential skills. *Discourse Processes, 48*, 119-138.

Freed, J., Lockton, E., & Adams, C. (2012). Short term and working memory skills in primary school aged children with specific language impairment and children with pragmatic language impairment: Phonological, linguistic and visuo spatial aspects. *International Journal of Language & Communication Disorders, 47,* 457-466. doi: 10.1111/j.1460-6984.2012.00148.x

Fujiki, M., Brinton, B., & Clarke, D. (2002). Emotion regulation in children with specific language impairment. *Language, Speech & Hearing Services in Schools, 33,* 102.

Garon, N., Bryson, S., & Smith, I. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin, 134*, 31-60. doi: 10.1037/0033-2909.134.1.31

Gathercole, S. E. (1995). Is non-word repetition a test of phonological memory or long-term knowledge? It all depends on the non-words. *Memory & cognition, 23,* 83-94.

Gathercole, S. E. (2006). Non-word repetition and word learning: The nature of the relationship. *Applied Psycholinguistics, 27*, 513-543.

Gathercole, S. E., & Adams, A. (1993). Phonological working memory in very young children. *Developmental Psychology, 29*, 770-778. doi: 10.1037/0012-1649.29.4.770

Gathercole, S. E., & Baddeley, A. (1990). Phonological memory deficits in language disordered children: Is there a causal connection? *Journal of Memory and Language, 29*, 336-360.

Gathercole, S. E., Tiffany, C., Briscoe, J., Thorn, A., & team, T. A. (2005). Developmental consequences of poor phonological short-term memory function in childhood: a longitudinal study. *Journal of Child Psychology and Psychiatry, 46*, 598-611. doi: 10.1111/j.1469-7610.2004.00379.x

Gathercole, S. E., Willis, C. S., Emslie, H., & Baddeley, A. D. (1992). Phonological memory and vocabulary development during the early school years: A longitudinal study. *Developmental Psychology, 28*, 887-898. doi: 10.1037/0012-1649.28.5.887

Gaulin, C. A., & Campbell, T. F. (1994). Procedure for assessing verbal working memory in normal school aged children; some preliminary data. *Perceptual and Motor Skills, 79,* 55-64.

Gillam, R. B., Cowan, N., & Day, L. S. (1995). Sequential memory in children with and without language impairment. *Journal of Speech & Hearing Research, 38*, 393-402.

Gioia, G., A., Epsy, K. A., & Isquith, P., K. . (2003). *Behaviour Rating Inventory of Executive Function - Preschool (BRIEF-P)* Lutz, FL: Psychological Assessment Resources.

Glascoe, F. P. (1997). *Parent Evaluation of Developmental Status: A method for detecting and addressing developmental and behavioural problems in children*. Nashville, Tennessee: Ellsworth & Vandermeer Press Ltd.

Gray, S. (2003). Diagnostic accuracy and test-retest reliability of non-word repetition and digit span tasks administered to preschool children with specific language impairment. *Journal of Communication Disorders, 26*, 129-151.

Griffith, E. M., Pennington, B. F., Wehner, E. A., & Rogers, S. J. (1999). Executive functions in young children with autism. *Child Development, 70*, 817-832.

Gupta, P., & Tisdale, J. (2009). Does phonological short-term memory causally determine vocabulary learning? Toward a computational resolution of the debate. *Journal of Memory and Language, 61*, 481-502. doi: 10.1016/j.jml.2009.08.001

Hart, B., & Risley, T. (1995). *Meaningful differences in the everyday experiences of young American children*. Baltimore, MD: Paul H. Brookes.

Hayiou-Thomas, M. E., Bishop, D., & Plunkett, K. (2004). Simulating SLI: General cognitive processing stressors can produce a specific linguistic profile. *Journal of Speech, Language and Hearing Research, 47*, 1347-1362. doi: 10.1044/1092-4388(2004/101)

Henrichs, J., Rescorla, L., Schenk, J. J., Schmidt, H. G., Jaddoe, V. W. V., Hofman, A., . . . Tiemeier, H. (2011). Examining continuity of early expressive vocabulary development: The Generation R study. *Journal of Speech, Language and Hearing Research, 54*, 854-869. doi: 10.1044/1092-4388(2010/09-0255)

Henry, L. A., Messer, D. J., & Nash, G. (2012). Executive functioning in children with specific language impairment. *Journal of Child Psychology and Psychiatry, 53*, 37-45. doi: 10.1111/j.1469-7610.2011.02430.x

Hill, E. L. (2001). Non-specific nature of specific language impairment: a review of the literature with regard to concomitant motor impairments. *International Journal of Language & Communication Disorders, 36*, 149-171. doi: 10.1080/13682820010019874

Hoff, E., Core, C., & Bridges, K. (2008). Non-word repetition assesses phonological memory and is related to vocabulary development in 20- to 24-month-olds. *Journal of Child Language, 35*, 903-916. doi: 10.1017/S0305000908008751

Horwitz, S. M., Irwin, J. R., Briggs-Gowan, M. J., Bosson Heenan, J. M., Mendoza, J., & Carter, A. S. (2003). Language delay in a community cohort of young children. *Journal of the American Academy of Child and Adolescent Psychiatry, 42*, 932-940. doi: 10.1097/01.chi.0000046889.27264.5e

Hughes, C. (1998). Finding your marbles: Does preschoolers' strategic behavior predict later understanding of mind? *Developmental Psychology, 34*, 1326-1339. doi: 10.1037/0012-1649.34.6.1326

Im-Bolter, N., Johnson, J., & Pascual-Leone, J. (2006). Processing limitations in children with specific language impairment: the role of executive function. *Child Development, 77*, 1822-1841.

Irwin, J. R., Carter, A. S., & Briggs-Gowan, M. J. (2002). The social-emotional development of “late-talking” toddlers. *Journal of the American Academy of Child & Adolescent Psychiatry, 41*, 1324-1332. doi: 10.1097/00004583-200211000-00014

Isquith, P. K., Crawford, J. S., Espy, K. A., & Gioia, G. A. (2005). Assessment of executive function in preschool-aged children. *Mental Retardation and Developmental Disabilities Research Reviews, 11*, 209-215. doi: 10.1002/mrdd.20075

Isquith, P. K., Gioia, G. A., & Espy, K. A. (2004). Executive function in preschool children: Examination through everyday behavior. *Developmental Neuropsychology, 26*, 403-422. doi: 10.1207/s15326942dn2601_3

Jacquemot, C., & Scott, S. K. (2006). What is the relationship between phonological short-term memory and speech processing? *Trends in Cognitive Sciences, 10*, 480-486. doi: 10.1016/j.tics.2006.09.002

Johnson, J., Fabian, V., & Pascual-Leone, J. (1989). Quantitative hardware stages that constrain language development. *Human development, 32*, 245-271. doi: 10.1159/000276477

Johnston, J. (1994). Cognitive abilities of children with language impairment. In M. Rice & R. Watkins (Eds.), *Specific language impairments in children* (pp. 107-120). Baltimore, MD: Brooks Publishing.

Just, M., & Carpenter, P. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*, 122-149. doi: 10.1037/0033-295x.99.1.122

Kamhi, A. G. (1995). Hierarchical planning abilities in children with Specific Language Impairments. *Journal of speech and hearing research, 38*, 1108-1116.

Kamhi, A. G., Catts, H., Koenig, L., & Lewis, B. (1984). Hypothesis-Testing and Nonlinguistic Symbolic Abilities in Language-Impaired Children. *Journal of Speech and Hearing Disorders, 49*, 169-176.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General*, *133*, 189-217. doi: 10.1037/0096-3445.133.2.189

Karmiloff-Smith, A. (2009). Nativism versus neuroconstructivism: Rethinking the study of developmental disorders. *Developmental Psychology, 45*, 56-63. doi: 10.1037/a0014506

Kidd, E. (2013). The role of verbal working memory in children's sentence comprehension: A critical review. *Topics in language disorders, 33*, 208-233.

Klee, T., Carson, D. K., Gavin, W. J., Hall, L., Kent, A., & Reece, S. (1998). Concurrent and predictive validity of an early language screening program. *Journal of Speech Language and Hearing Research, 41*, 627-641.

Kronenberger, W. G., Pisoni, D. B., Harris, M. S., Hoen, H. M., Xu, H., & Miyamoto, R. T. (2013). Profiles of verbal working memory growth predict speech and language development in children with cochlear implants. *Journal of Speech, Language, and Hearing Research, 56*, 805-825. doi: 10.1044/1092-4388(2012/11-0356)

La Paro, K. M., Justice, L., Skibbe, L. E., & Pianta, R. C. (2004). Relations among maternal, child, and demographic factors and the persistence of preschool language impairment. *American Journal of Speech - Language Pathology, 13*, 291-303. doi: 10. 1111/1467-8624.00313.2001-01005-01010.1111/1467-8624.00313

Law, J., Garrett, Z. & Nye, C. (2004). The efficacy of treatment for children with developmental delay / disorder: A meta-analysis. *Journal of Speech, Language and Hearing Research*, *47*, 924-943.

Leonard, L. B. (1998). *Children with specific language impairment*. Cambridge, Massachusetts: Massachusetts Institute of Technology.

Leonard, L. B., Ellis Weismer, S., Miller, C., Francis, D., Tomblin, J. B., & Kail, R. (2007). Speed of processing, working memory, and language impairment in children. *Journal* *of Speech, Language & Hearing Research, 50*, 408-428. doi: 10.1044/1092-4388(2007/029)

Lum, J. A. G., Conti-Ramsden, G., Page, D., & Ullman, M. T. (2012). Working, declarative and procedural memory in specific language impairment. *Cortex, 48*, 1138-1154. doi: 10.1016/j.cortex.2011.06.001

Lyytinen, P., Poikkeus, A.-M., Laakso, M.-L., Eklund, K., & Lyytinen, H. (2001). Language development and symbolic play in children with and without familial risk for dyslexia. *Journal of Speech, Language, and Hearing Research, 44*, 873-885. doi: 10.1044/1092-4388(2001/070)

Maatta, S., Laakso, M.-L., Tolvanen, A., Ahonim, T., & Ara, T. (2012). Developmental trajectories of early communication skills. *Journal of Speech, Language & Hearing Research, 55*, 1083-1096. doi: 10.1044/1092-4388(2011/10-0305)

MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review, 109*(1), 35-54. doi: 10.1037/0033-295x.109.1.35

MacRoy-Higgins, M., & Kaufman, I. (2012). Pragmatic functions of toddlers who are late talkers. *Communication Disorders Quarterly, 33*, 242-244. doi: 10.1177/1525740112447942

MacRoy-Higgins, M., Schwartz, R. G., Shafer, V. L., & Marton, K. (2013). Influence of phonotactic probability/neighbourhood density on lexical learning in late talkers. *International Journal of Language & Communication Disorders, 48*, 188-199. doi: 10.1111/j.1460-6984.2012.00198.x

Magimairaj, B. M., & Montgomery, J. W. (2012a). Children's verbal working memory: Relative importance of storage, general processing speed, and domain-general controlled attention. *Acta Psychologica, 140*, 196-207. doi: 10.1016/j.actpsy.2012.05.004

Magimairaj, B. M., & Montgomery, J. W. (2012b). Children's verbal working memory: Role of processing complexity in predicting spoken sentence comprehension. *Journal of Speech, Language, and Hearing Research, 55*, 669-682. doi: 10.1044/1092-4388(2011/11-0111)

Mainela Arnold, E., Misra, M., Miller, C., Poll, G. H., & Park, J. S. (2012). Investigating sentence processing and language segmentation in explaining children's performance on a sentence span task. *International Journal of Language & Communication Disorders, 47*, 166-175. doi: 10.1111/j.1460-6984.2011.00080.x

Marchman, V. A., & Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental Science, 11*, 9-16. doi: 10.1111/j.1467-7687.2008.00671.x

Marcovitch, S., & Zelazo, P. D. (1999). The A-not-B error: Results from a logistic meta-analysis. *Child Development, 70*, 1297-1313. doi: 10.2307/1132307

Martin, N., & Brownell, R. (2011). *The Receptive One Word Picture Vocabulary Test (Fourth Edition)* (Fourth ed.). Novato, CA: Academic Therapy Publications.

Marton, K. (2008). Visuo-spatial processing and executive functions in children with Specific Language Impairment. *International Journal of Language & Communication Disorders, 43*, 181-200. doi: 10.1080/16066350701340719

Marton, K., Campanelli, L., Scheuer, J., Yoon, J., & Eichorn, N. (2012). Executive function profiles in children with and without Specific Language Impairment. *Rivista Di Psicolinguistica Applicata, 12*, 57-73.

Marton, K., & Schwartz, R. (2003). Working memory capacity and language processes in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 46*, 1138-1153.

Max Planck Institute for Psycholinguistics, The Language Archives. (2012). ELAN (Version 4.4.0). Nijmegen, The Netherlands. Retrieved from http://tla.mpi.nl/tools/tla-tools/elan/

McCabe, P. (2005). Social and behavioral correlates of preschoolers with Specific Language Impairment. *Psychology in the Schools, 42*, 373-387. doi: 10.1002/pits.20064

McIntosh, B., & Dodd, B. (2011). *Toddler Phonology Test*. London, UK: Pearson Assessment.

Melby-Lervag, M., Lervag, A., Lyster, S. A., Klem, M., Hagtvet, B., & Hulme, C. (2012). Nonword-repetition ability does not appear to be a causal influence on children's vocabulary development. *Psychological Science, 23*, 1092-1098. doi: 10.1177/0956797612443833

Metsala, J. L. (1999). Young children's phonological awareness and nonword repetition as a function of vocabulary development. *Journal of Educational Psychology, 91*, 3-19. doi: 10.1037/0022-0663.91.1.3

Miles, J., & Shevlin, M. (2001). *Applying regression and correlation: A guide for students and researchers*. London: Sage Publications.

Miller, C., Kail, R., Leonard, L., & Tomblin, B. (2001). Speed of processing in children with specific language impairment. *Journal of Speech Language and Hearing Research, 44,* 416-433. doi: 10.1044/1092-4388(2001/034)

Miller, C., Leonard, L. B., Kail, R. V., Zhang, X., Tomblin, J. B., & Francis, D. J. (2006). Response time in 14-year-olds with language impairment. *Journal of Speech, Language, and Hearing Research, 49*, 712-728. doi: 10.1044/1092-4388(2006/052)

Miller, M. R., Giesbrecht, G. F., Müller, U., McInerney, R. J., & Kerns, K. A. (2012). A latent variable approach to determining the structure of executive function in preschool children. *Journal of Cognition and Development, 13*, 395-423. doi: 10.1080/15248372.2011.585478

Mirak, J., & Rescorla, L. (1998). Phonetic skills and vocabulary size in late talkers: Concurrent and predictive relationships. *Applied Psycholinguistics, 19*, 1-17. doi: 10.1017/S0142716400010559

Miyake, A., Friedman, N., Emerson, M., Witzki, A., Howerter, A., & Wager, T. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*, 49-100. doi: 10.1006/cogp.1999.0734

Montgomery, J. (1993). Haptic recognition of children with specific language impairment: Effects of response modality. *Journal of Speech and Hearing Research, 36,* 98-104.

Montgomery, J. (2000a). Relation of working memory to off-line and real-time sentence processing in children with specific language impairment. *Applied Psycholinguistics, 21*(01), 117-148.

Montgomery, J. (2000b). Verbal working memory and sentence comprehension in children with specific language impairment. *Journal of Speech, Language, Hearing Research, 43*, 293-308.

Montgomery, J., & Evans, J. (2009). Complex sentence comprehension and working memory in children with specific language impairment. *Journal of Speech, Language and Hearing Research, 52,* 269-288. doi: 10.1044/1092-4388(2008/07-0116)

Montgomery, J., Magimairaj, B., & Finney, M. (2010). Working memory and specific language impairment: An update on the relation and perspectives on assessment and treatment. *American Journal of Speech-Language Pathology, 19*, 78-94. doi: 10.1044/1058-0360(2009/09-0028)

Montgomery, J., Polenenko, A., & Marienellie, S. (2009). Role of working memory in children's understanding spoken narrative: A preliminary investigation. *Applied Psycholinguistics, 30*, 485-509. doi: 10.1017/S0142716409090249

Morrison, C. M., Chappell, T. D., & Ellis, A. W. (1997). Age of acquisition norms for a large set of object names and their relation to adult estimates and other variables. *The Quarterly Journal of Experimental Psychology Section A, 50*, 528-559. doi: 10.1080/027249897392017

Moyle, J., Stokes, S. F., & Klee, T. (2011). Early language delay and specific language impairment. *Developmental Disabilities Research Reviews, 17*, 160-169. doi: 10.1002/ddrr.1110

Mullen, E. (1995). *Mullen Scales of Early Learning (AGS Edition)*. San Antonio, TX: Pearson Clinical Assessment.

Munroe, N., Baker, E., McGregor, K., Docking, K., & Arculi, J. (2012). Why word learning is not fast. *Frontiers in Psychology, 3*(41), 1-10.

Nairne, J. S. (2002). Remembering over the short-term: The case against the standard model. *Annual Review of Psychology, 53*, 53-81.

Nee, D. E., Brown, J. W., Askren, M. K., Berman, M. G., Demiralp, E., Krawitz, A., & Jonides, J. (2013). A meta-analysis of executive components of working memory. *Cerebral Cortex, 23*, 264-282. doi: 10.1093/cercor/bhs007

Newman, S., Malaia, E., Seo, R., & Cheng, H. (2013). The effect of individual differences in working memory capacity on sentence comprehension: An fMRI study. *Brain Topography, 26*, 458-467. doi: 10.1007/s10548-012-0264-8

Noterdaeme, M., Amorosa, H., Mildenberger, K., Sitter, S., & Minow, F. (2001). Evaluation of attention problems in children with autism and children with a specific language disorder. *European Child & Adolescent Psychiatry, 10*, 58-66.

Oberauer, K. (2009). Design for a working memory. *Psychology of Learning and Motivation* (Vol. 51, pp. 45-100). San Diego: Elsevier Academic Press Inc. doi: 10.1016/S0079-7421(09)51002-X

Olswang, L., B., Rodriguez, B and Timler, G. (1998). Recommending intervention for toddlers with specific language learning difficulties: We may not have all the answers, but we know a lot. *Journal of Speech-Language Pathology, 7*, 23-32.

Oram Cardy, J. E., Flagg, E. J., Roberts, W., & Roberts, T. P. L. (2008). Auditory evoked fields predict language ability and impairment in children. *International Journal of Psychophysiology, 68*, 170-175. doi: 10.1016/j.ijpsycho.2007.10.015

Oram, J. E. (2003). *Seeking cognitive markers of specific language impairment and attention-deficit/hyperactivity disorder.* Doctoral dissertation. Retrieved from ProQuest LLC. (305256355)

Paul, R. (1991). Profiles of toddlers with slow expressive language development. *Topics in Language Disorders, 11*, 1-13.

Paul, R. (1993). Patterns of development in late talkers: Preschool years. *Communication Disorders Quarterly, 15*, 7-14. doi: 10.1177/152574019301500103

Paul, R. (1996). Clinical implications of the natural history of slow expressive language development. *American Journal of Speech Language Pathology, 5*(2), 5-21.

Paul, R. (2007). Assessment and intervention for emerging language. In R. Paul (Ed.), *Language disorders from infancy through adolescence* (Third Edition ed., pp. 264-316). St Louis, Missouri: Mosby, Elsevier.

Paul, R., & Alforde, S. (1993). Grammatical morpheme acquisition in four-year-olds with normal, impaired, and late-developing language. *Journal of Speech & Hearing Research, 36*, 1271-1275.

Paul, R., & Elwood, T. J. (1991). Maternal linguistic input to toddlers with slow expressive language development. *Journal of Speech & Hearing Research, 34*, 982-988.

Paul, R., Hernandez, R., Taylor, L., & Johnson, K. (1996). Narrative development in late Talkers: Early school age. *Journal of Speech and Hearing Research, 39*, 1295-1303.

Paul, R., Looney, S. S., & Dahm, P. S. (1991). Communication and socialization skills at ages 2 and 3 in "late-talking" young children. *Journal of Speech & Hearing Research, 34*(4), 858-865.

Paul, R., & Roth, F. P. (2011). Characterizing and predicting outcomes of communication delays in infants and toddlers: Implications for clinical practice. *Language, Speech & Hearing Services in Schools, 42,* 331-340. doi: 10.1044/0161-1461(2010/09-0067)

Paul, R., & Shiffer, M. E. (1991). Communication initiations in normal and late talking toddlers. *Applied Psycholinguistics, 12*, 419-431.

Petruccelli, N., Bavin, E. L., & Bretherton, L. (2012). Children with specific language impairment and resolved late talkers: Working memory profiles at 5 years. *Journal of Speech, Language, and Hearing Research, 55*, 1690-1703. doi: 10.1044/1092-4388(2012/11-0288)

Piaget, J. (1954). *The construction of reality in the child*. London: Basic Books.

Pierpont, E. I., Richmond, E. K., Abbeduto, L., Kover, S. T., & Brown, W. T. (2011). Contributions of phonological and verbal working memory to language development in adolescents with fragile X syndrome. *Journal of Neurodevelopmental Disorders, 3*, 335-347. doi: 10.1007/s11689-011-9095-2

Pisoni, D. B., Kronenberger, W. G., Roman, A. S., & Geers, A. E. (2011). Measures of digit span and verbal rehearsal speed in Deaf children after more than 10 years of cochlear implantation. *Ear and Hearing, 32*(1 Suppl, Long-Term Outcomes of Cochlear Implantation in Early Childhood), 60S-74S. doi: 10.1097/AUD.0b013e3181ffd58e

Poll, G. H., & Miller, C. A. (2013). Late talking, typical talking, and weak language skills at middle childhood. *Learning and Individual Differences, 26*, 177-184. doi: 10.1016/j.lindif.2013.01.008

Poll, G. H., Miller, C. A., Mainela-Arnold, E., Adams, K. D., Misra, M., & Park, J. S. (2013). Effects of children's working memory capacity and processing speed on their sentence imitation performance. *International Journal of Language & Communication Disorders, 48*, 329-342. doi: 10.1111/1460-6984.12014

Preston, J. L., Frost, S. J., Mencl, W. E., Fulbright, R. K., Landi, N., Grigorenko, E., . . . Pugh, K. R. (2010). Early and late talkers: School-age language, literacy and neurolinguistic differences. *Brain: A Journal of Neurology, 133*, 2185-2195. doi: 10.1093/brain/awq163

Price, T. S., Eley, T. C., Dale, P. S., Stevenson, J., Saudino, K., & Plomin, R. (2000). Genetic and environmental covariation between verbal and nonverbal cognitive development in infancy. *Child Development, 71*, 948-959.

Psychology Software Tools Inc. (2010). E-Prime Professional Studio (Version 2.0.8.90). Sharpsberg, PA, USA.

Psychology Software Tools Inc. (2012). E-Prime Extensions for Tobii (2.0.1.26). Sharpsberg, PA, USA.

Redmond, S. M., & Rice, M. L. (1998). The socioemotional behaviors of children with SLI: Social adaptation or social deviance? *Journal of Speech, Language, and Hearing Research, 41*, 688-700.

Reese, E., & Read, S. (2000). Predictive validity of the New Zealand MacArthur Communicative Development Inventory: Words and Sentences. *Journal of Child Language, 27*(02), 255-266.

Reilly, S., Wake, M., Bavin, E. L., Prior, M., Williams, J., Bretherton, L., . . . Ukoumunne, O. C. (2007). Predicting language at two years of age: A prospective community study. *Pediatrics, 120*, e1441-e1449. doi: 10.1542/peds.2007-0045

Reilly, S., Wake, M., Ukoumunne, O. C., Bavin, E., Prior, M., Cini, E., . . . Bretherton, L. (2010). Predicting language outcomes at four years of age: Findings from early language in Victoria study. *Pediatrics, 126*, e1530-e1537. doi: 10.1542/peds.2010-0254

Rescorla, L. (1989). The Language Development Survey: A screening tool for delayed language in toddlers. *The Journal of Speech and Hearing Disorders, 54*, 587-599.

Rescorla, L. (2000). Do late-talking toddlers turn out to have reading difficulties a decade later? *Annals of Dyslexia, 50*, 87-102.

Rescorla, L. (2009). Age 17 language and reading outcomes in late-talking toddlers: Support for a dimensional perspective on language delay. *Journal of Speech, Language & Hearing Research, 52*, 16-30.

Rescorla, L. (2011). Late talkers: Do good predictors of outcome exist? *Developmental Disabilities Research Reviews, 17*, 141-150. doi: 10.1002/ddrr.1108

Rescorla, L., Bascome, A., Lampard, J., & Feeny, N. (2001). Conversational patterns in late talkers at age 3. *Applied Psycholinguistics, 22*, 235-251.

Rescorla, L., Dahlsgaard, K., & Roberts, J. (2000). Late-talking toddlers: MLU and IPSyn outcomes at 3;0 and 4;0. *Journal of Child Language, 27*, 643-664. doi: 10.1017/S0305000900004232

Rescorla, L., & Merrin, L. (1998). Communicative intent in late-talking toddlers. *Applied Psycholinguistics, 19*, 393-414.

Rescorla, L., & Roberts, J. (1997). Late talkers at 2: Outcome at age 3. *Journal of Speech, Language & Hearing Research, 40*, 556-566.

Reynell, J. K. (1977). *Reynell Developmental Language Scales*. Windsor: NFER.

Rheinheimer, D. C., & Penfield, D. A. (2001). The effects of type I error rate and power of the ANCOVA F test and selected alternatives under non-normality and variance heterogeneity. *The Journal of Experimental Education, 69*, 373-391.

Rice, M. L., Taylor, C., & Zubrick, S. (2008). Language outcomes of seven-year-old children with or without a history of late language emergence at 24 months. *Journal of Speech, Language & Hearing Research, 51*, 394-407.

Rice, M. L., Wexler, K., & Cleave, P. (1995). Specific language impairment as a period of extended optional infinitive. *Journal of Speech Hearing Research, 38*, 850-863.

Roberts, J., Rescorla, L., Giroux, J., & Stevens, L. (1998). Phonological skills of children with specific expressive language impairment (SLI-E): Outcome at age three. *Journal of Speech, Language & Hearing Research, 41*, 374-384.

Roulstone, S., Peters, T. J., Glogowska, M., & Enderby, P. (2003). A 12-month follow-up of preschool children investigating the natural history of speech and language delay. *Child: Care, Health and Development, 29*, 245-255. doi: 10.1046/j.1365-2214.2003.00339.x

Rowe, M. L., Raudenbush, S. W., & Goldin-Meadow, S. (2012). The pace of vocabulary growth helps predict later vocabulary skill. *Child Development, 83*, 508-525. doi: 10.1111/j.1467-8624.2011.01710.x

Roy, P. & Chiat, S. (2004). A prosodically controlled word and non-word repetition task for 2- to 4- year-olds: Evidence from typically developing children. *Journal of Speech, Language and Hearing Research, 47*, 223-234.

Saffran, J. R., & Wilson, D. P. (2003). From syllables to syntax: Multilevel statistical learning by 12-month-old infants. *Infancy, 4*, 273-284. doi: 10.1207/S15327078IN0402_07

Schafer, J. L. & Graham, J. W. (2002). *Missing data: Our view of the state of the art. Psychological methods, 7*, 147-177.

Scarborough, H. S. (1990). Index of Productive Syntax. *Applied Psycholinguistics, 11*, 1-12.

Schjølberg, S., Eadie, P., Zachrisson, H. D., Oyen, A.-S., & Prior, M. (2011). Predicting language development at age 18 months: data from the Norwegian Mother and Child Cohort Study. *Journal of Developmental and Behavioral Pediatrics, 32*, 375-383. doi: 10.1097/DBP.0b013e31821bd1dd

Schuchardt, K., Bockmann, A.-K., Bornemann, G., & Maehler, C. (2013). Working memory functioning in children with learning disorders and specific language impairment. *Topics in Language Disorders, 33*, 298-312. doi: 10.1097/01.TLD.0000437943.41140.36

Semel, E., Wiig, E. H., & Secord, W. A. (2003). *Clinical Evaluation of Language Fundamentals - Fourth Edition (CELF-4)* (Fourth Edition ed.). San Antonio, TX: Harcourt Assessment.

Senn, T., Espy, K. A., & Kaufmann, P. (2004). Using path analysis to understand executive function organization in preschool children. *Developmental Neuropsychology, 26*, 445-464. doi: 10.1207/s15326942dn2601_5

Silva, P. A., Williams, S., & McGee, R. (1987). A longitudinal study of children with developmental language delay at age three: Later intelligence, reading and behaviour problems. *Developmental Medicine & Child Neurology, 29*, 630-640. doi: 10.1111/j.1469-8749.1987.tb08505.x

Silvestri, R. (Writer). (2009). Ormie. In A. Productions (Producer). Toronto, Ontario.

Smith, L. (1998). Predicting communicative competence at 2 and 3 years from pragmatic skills at 10 months. *International Journal of Language & Communication Disorders, 33*, 127-148. doi:10.1080/136828298247794

Smith, L., & Ulvund, S. E. (1998). The relation of joint attention at 13 months to subsequent language, verbal IQ and nonverbal IQ measures through middle childhood. *Infant Behavior and Development, 21*(Supplement 1), 142-142. doi: 10.1016/S0163-6383(98)91357-1

Spaulding, T. (2010). Investigating mechanisms of suppression in preschool children with specific language impairment. *Journal of Speech, Language & Hearing Research, 53*, 725-738. doi: 10.1044/1092-4388(2009/09-0041)

Stanton-Chapman, T. L., Chapman, D. A., Bainbridge, N. L., & Scott, K. G. (2002). Identification of early risk factors for language impairment. *Research in Developmental Disabilities, 23*, 390-405. doi: http://dx.doi.org/10.1016/S0891-4222(02)00141-5

StatisticsNZ. (2006). *New Zealand Census Information.* Retrieved from http://www.stats.govt.nz/Census/2006CensusHomePage/QuickStats.aspx

Stokes, S. F. (2010). Neighborhood density and word frequency predict vocabulary size in toddlers. *Journal of Speech Language and Hearing Research, 53*, 670-683. doi: 10.1044/1092-4388(2009/08-0254)

Stokes, S. F. (2013). The impact of phonological neighborhood density on typical and atypical emerging lexicons. *Journal of Child Language, 41*, 634-657. doi: 10.1017/S030500091300010X

Stokes, S. F., Kern, S., & Dos Santos, C. (2012). Extended statistical learning as an account for slow vocabulary growth. *Journal of Child Language, 39*, 105-129. doi: 10.1017/S0305000911000031

Stokes, S. F., & Klee, T. (2009a). The diagnostic accuracy of a new test of early nonword repetition for differentiating late talking and typically developing children. *Journal of Speech Language and Hearing Research, 52,* 872-882. doi: 10.1044/1092-4388(2009/08-0030)

Stokes, S. F., & Klee, T. (2009b). Factors that influence vocabulary development in two-year-old children. *Journal of Child Psychology and Psychiatry, 50,* 498-505. doi: 10.1111/j.1469-7610.2008.01991.x

Stothard, S. E., Snowling, M. J., Bishop, D. V. M., Chipchase, B. B., & Kaplan, C. A. (1998). Language-impaired preschoolers: A follow-up into adolescence. *Journal of Speech, Language, and Hearing Research, 41,* 407-418.

Sun, R., Zhang, X., Slusarz, P., & Mathews, R. (2007). The interaction of implicit learning, explicit hypothesis testing learning and implicit-to-explicit knowledge extraction. *Neural Networks, 20*, 34-47. doi: 10.1016/j.neunet.2006.07.002

Swingley, D., Pinto, J. P., & Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognition, 71*(2), 73-108. doi: 10.1016/s0010-0277(99)00021-9

Tallal, P., & Stark, R. (1981). Speech acoustic-cue discrimination abilities of normally developing and language-impaired children. *Journal of the Acoustical Society of America, 69*, 568-574.

Taylor, C. L., Christensen, D., Lawrence, D., Mitrou, F., & Zubrick, S. R. (2013). Risk factors for children’s receptive vocabulary development from four to eight years in the longitudinal study of Australian children. *PLOS ONE, 8*(9), 1-20. doi: 10.1371/journal.pone.0073046

Thal, D., Miller, S., Carlson, J., & Vega, M. M. (2005). Nonword repetition and language development in 4-year-old children with and without a history of early language delay.

Journal of Speech Language and Hearing Research, 48, 1481-1495. doi: 10.1044/1092-4388(2005/103)

Thal, D., Oroz, M., & McCaw, V. (1995). Phonological and lexical development in normal and late-talking toddlers. *Applied Psycholinguistics, 16*, 407-424.

Thal, D., Reilly, J., Seibert, L., Jeffries, R., & Fenson, J. (2004). Language development in children at risk for language impairment: Cross-population comparisons. *Brain and Language, 88*, 167-179. doi: 10.1016/s0093-934x(03)00096-8

Thal, D., Tobias, S., & Morrison, D. (1991). Language and gesture in late talkers: a one-year follow-up. *Journal of Speech Hearing Research, 34*, 604-612.

Tobii Technology (2012). Tobii Studio (Version 3.0.3.270). Stockholm, Sweden.

Tomblin, J. B., Hardy, J. C., & Hein, H. A. (1991). Predicting poor-communication status in preschool children using risk factors present at birth. *Journal of Speech & Hearing Research, 34*, 1096-1105.

Tomblin, J. B., Records, N., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *Journal of Speech, Language & Hearing Research, 40*, 1245-1260.

Tomblin, J. B., & Zhang, X. (1999). Are children with specific language impairment a unique group of language learners? In H. Tager-Flusberg (Ed.), *Neurdevelopmental Disorders: Contributions to a new framework from the cognitive neurosciences* (pp. 361-382). Cambridge, MA: MIT Press.

Turkheimer, E., Haley, A., Waldron, M., D'Onofrio, B., & Gottesman, I. I. (2003). Socioeconomic status modifies heritability of IQ in young children. *Psychological Science, 14*, 623-628.

Ullman, M. T. (2001). A neurocognitive perspective on language: The declarative/procedural model. *National Review Neuroscience, 2*, 717-726. doi: 10.1038/35094573

Ullman, M. T., & Pierpont, E. (2005). Specific language impairment is not specific to language: the procedural deficit hypothesis. *Cortex, 41*, 399-433. doi: 10.1016/s0010-9452(08)70276-4

van der Lely, H. (2005). Domain-specific cognitive systems: insight from Grammatical-SLI. *Trends in Cognitive Sciences, 9*(2), 53-59. doi: 10.1016/j.tics.2004.12.002

Vernes, S. C. D., Newbury, D. F. D., Abrahams, B. S. P., Winchester, L., Nicod, J., Groszer, M., . . . Fisher, S. E. D. (2008). A functional genetic link between distinct developmental language disorders. *The New England Journal of Medicine*, *359*, 2337-2345.

Vernon-Feagans, L., Garrett-Peters, P., Willoughby, M., & Mills-Koonce, R. (2011). Chaos, poverty, and parenting: Predictors of early language development. *Early Childhood Research Quarterly*, *27*, 339-351 doi: 10.1016/j.ecresq.2011.11.001

Vig, S., Kaminer, R. K., & Jedrysek, E. (1987). A later look at borderline and mildly retarded preschoolers. *Journal of Developmental and Behavioral Pediatrics, 8*, 12-17. doi: 10.1097/00004703-198702000-00004

Vigil, D. C., Hodges, J., & Klee, T. (2005). Quantity and quality of parental language input to late-talking toddlers during play. *Child Language Teaching & Therapy, 21*(2), 107-122.

Viterbori, P., Gandolfi, E., & Usai, M. C. (2012). Executive skills and early language development. *Rivista Di Psicolinguistica Applicata, 12*(3), 17-32.

Waters, G. S., & Caplan, D. (1996). The capacity theory of sentence comprehension: Critique of Just and Carpenter (1992). *Psychological Review*, *103*, 761-772. doi: 10.1037/0033-295x.103.4.761

Waters, G. S., Caplan, D., & Rochon, E. (1995). Processing capacity and sentence comprehension in patients with Alzheimer's disease. *Cognitive Neuropsychology, 12*, 1-30. doi: 10.1080/02643299508251990

Webster, R. I., Majnemer, A., Platt, R. W., & Shevell, M. I. (2004). The predictive value of a preschool diagnosis of developmental language impairment. *Neurology, 63*, 2327-2331.

Westerlund, M., Berglund, E., & Eriksson, M. (2006). Can severely language delayed 3-year-olds be identified at 18 Months? Evaluation of a screening version of the MacArthur-Bates Communicative Development Inventories. *Journal of Speech, Language and Hearing Research, 49,* 237-247. doi: 10.1044/1092-4388(2006/020)

Weyandt, L. L., & Willis, W. G. (1994). Executive functions in school-aged children: Potential efficacy of tasks in discriminating clinical groups. *Developmental Neuropsychology, 10,* 27-38. doi: 10.1080/87565649409540564

Whitehouse, A. J. O., Bishop, D. V. M., Ang, Q. W., Pennell, C. E., & Fisher, S. E. (2011). CNTNAP2 variants affect early language development in the general population. *Genes, brain, and behavior, 10,* 451-456. doi: 10.1111/j.1601-183X.2011.00684.x

Whitehurst, G. J., & Fischel, J. (1994). Practitioner review: Early developmental language delay: What, if anything, should the clinician do about it? *Journal of Child Psychology, Psychiatry, and Allied Disciplines, 35*, 613-648.

Whitehurst, G. J., Smith, M., Fischel, J. E., Arnold, D. S., & Lonigan, C. J. (1991). The continuity of babble and speech in children with specific expressive language delay. *Journal of Speech Hearing Research, 34,* 1121-1129.

Wiebe, S., Espy, K. A., & Charak, D. (2008). Using confirmatory factor analysis to understand executive control in preschool children: I. Latent structure. *Developmental Psychology, 44*, 575-587. doi: 10.1037/0012-1649.44.2.575

Williams, L. A., & Elbert, M. (2003). A prospective longitudinal study of phonological development in late talkers. *Language, Speech & Hearing Services in Schools, 34*, 138-153. doi: 10.1044/0161-1461(2003/012)

Willoughby, M., Blair, C., Wirth, R. J., & Greenberg, M. (2010). The measurement of executive function at age 3 years: Psychometric properties and criterion validity of a new battery of tasks. *Psychological Assessment, 22,* 306-317. doi: 10.1037/a0018708

Wittke, K., Spaulding, T. J., & Schechtman, C. J. (2013). Specific language impairment and executive functioning: Parent and teacher ratings of behavior. *American Journal of Speech Language Pathology, 22*, 161-172. doi: 10.1044/1058-0360(2012/11-0052)

Zambrana, I. M., Ystrom, E., Schjølberg, S., & Pons, F. (2013). Action imitation at 1½ years is better than pointing gesture in predicting late development of language production at 3 years of age. *Child Development, 84,* 560-573. doi: 10.1111/j.1467-8624.2012.01872.x

Zaretsky, E. (2004). *Auditory comprehension in children with specific language impairment: The role of verbal working memory.* (Doctoral Dissertation). Retrieved from ProQuest LLC. (3101103)

Zimmerman, I., Lee, Steiner, V., G., & Pond, R., Evatt. (2002). *Preschool Language Scale (Fourth Edition)*. San Antonio, TX: Pearson.

Zubrick, S., Taylor, C., Rice, M., & Slegers, D. (2007). Late language emergence at 24 months: An epidemiological study of prevalence, predictors, and covariates. *Journal of Speech, Language & Hearing Research, 50*, 1562-1592. doi: 10.1044/1092-4388(2007/106)